DXOMARK

FOCUSED ON EXCELLENCE

MVA 2022-2023: The machine Intelligence of images – IA from cameras to IQ evaluation

IA from cameras to IQ evaluation

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

1. AI and Deep Learning

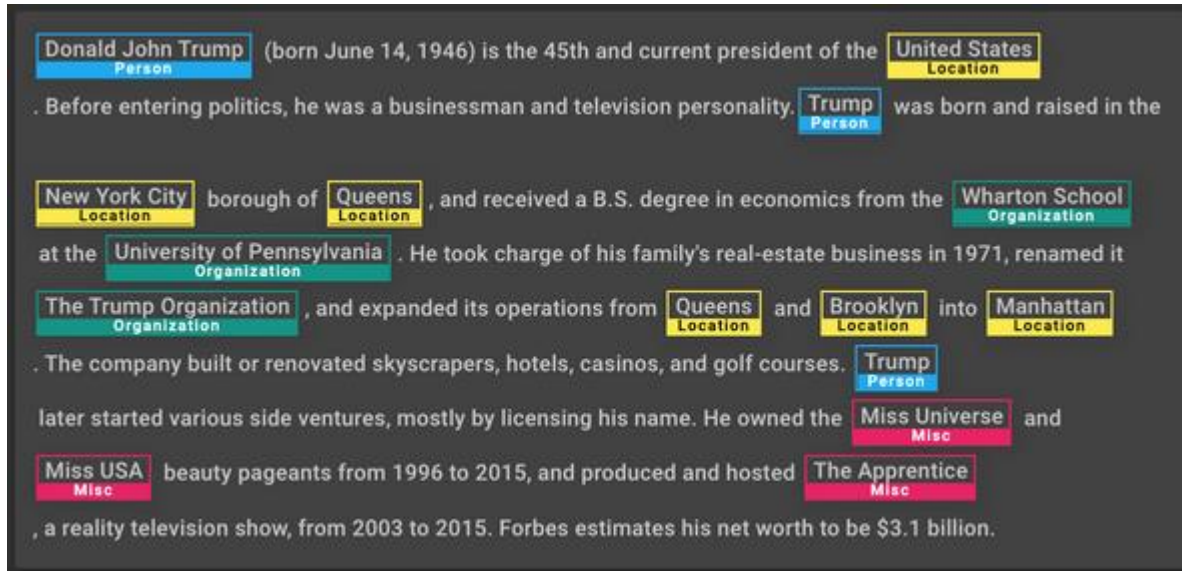
What do we mean by Artificial Intelligence?

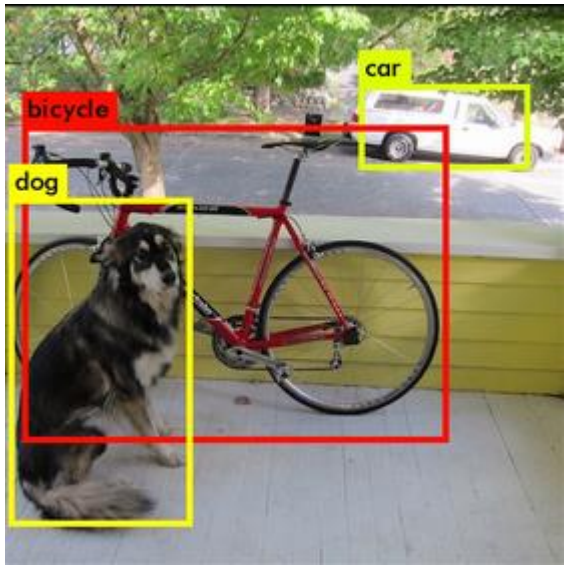
- Artificial Intelligence (AI) is an area of knowledge in the intersection between computer science, applied mathematics and perception. Its goal is to allow computers take actions depending on their environment and a given predefined goal.

- Within AI, we find wide areas of research such as Computer Vision or Natural language processing, which are interested in interpreting their environment through images and language, respectively

- The origin of AI backs to the 1950s.

- Currently, industrial solutions are widely dominated by “Machine Learning” and more specifically by Deep Learning.





https://doccano.herokuapp.com/fr

Emergence of Deep Learning

[2000s] Fast increase in computational power: GPUs (Graphic processing units) are processors optimized, at first, for graphics manipulation (rendering), very important in the gaming industry. Nvidia's CUDA plateform (2007) was the first widely adopted programming model for GPU computing. The main advantage of GPUs is their optimization for vectorial computations that can make them orders of magnitud faster that CPUs

[2009] Large scale Data bases such as ImageNet: Fei-Fei Li and colleagues in 2009 presented at CVPR with 14 million images hand-annotated (using crowdsourcing) 20k categories, 1M have also bounding boxes

[2012] Algorithmic solutions: 30 September 2012, a convolutional neural network (CNN) called AlexNet achieved a top-5 error of 15.3% in the ImageNet 2012 Challenge, more than 10.8 percentage points lower than that of the runner up. Their implementation on GPUs allowed to train on the complete database much faster than a CPU would

https://www.image-net.org/

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, **ImageNet: A Large-Scale Hierarchical Image Database.** *IEEE Computer Vision and Pattern Recognition (CVPR), 2009*

2022: Deep Learning-based models are now everywhere

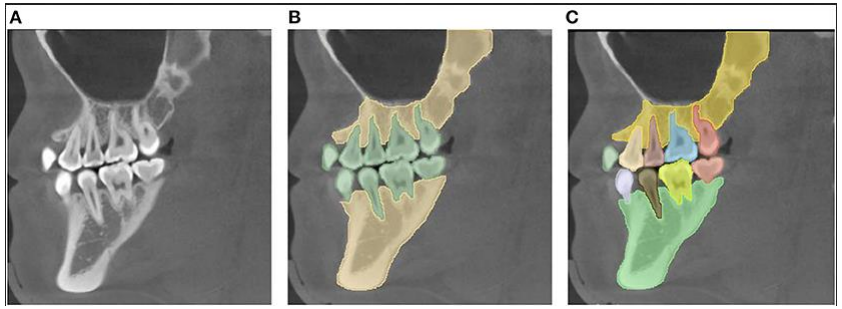
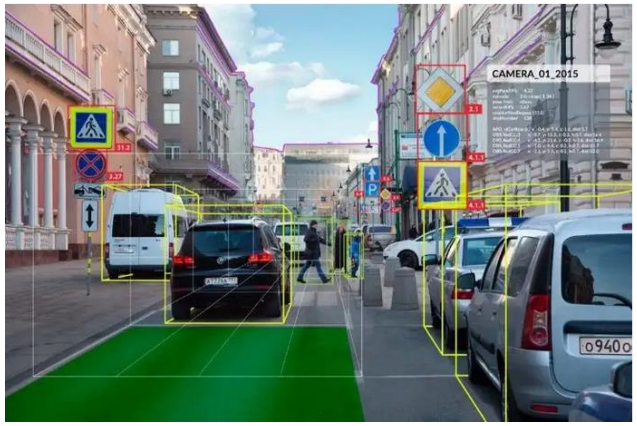


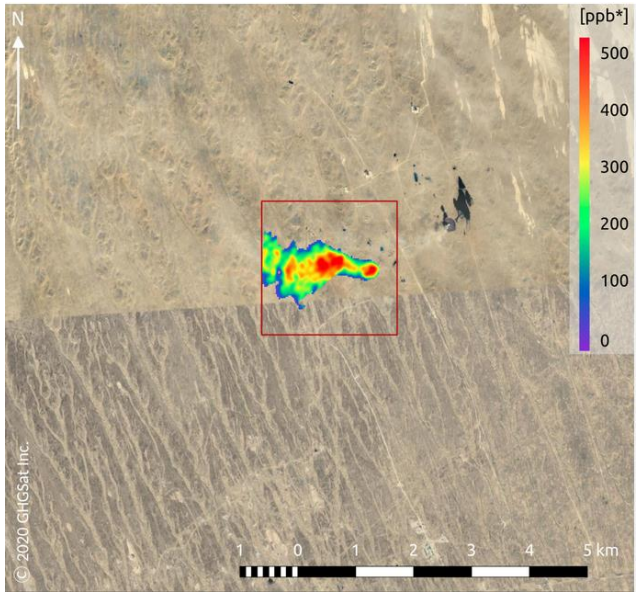
Image segmentation on X-Rays for clinical applications

https://www.frontiersin.org/articles/10.3389/fmedt.2021.767836/full



Self-driving cars

https://becominghuman.ai/computer-vision-applications-in-self-driving-cars-610561e14118

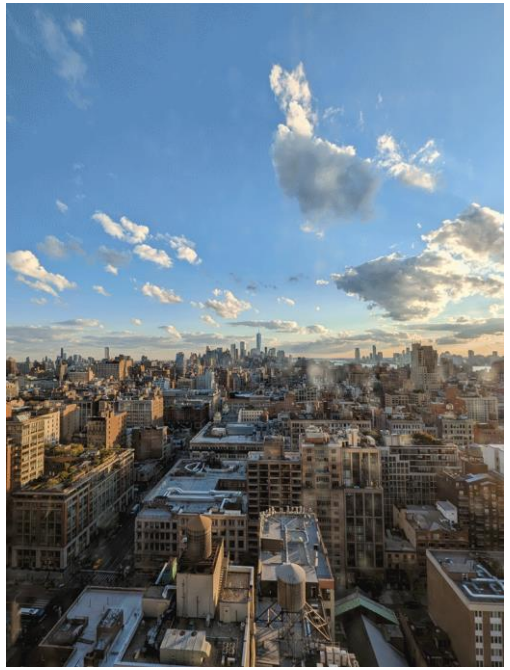


Satellite imagining for detection of gas leaks, detection

https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5P/Monitoring_methane_emissions_from_gas_pipelines

2022: Deep Learning-based models are now everywhere

Smartphone cameras are no exception



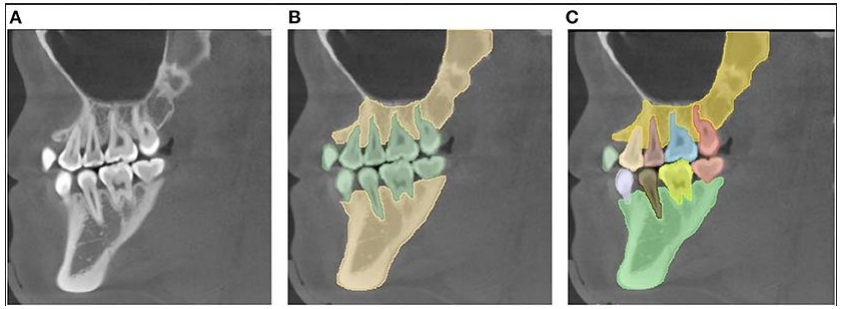
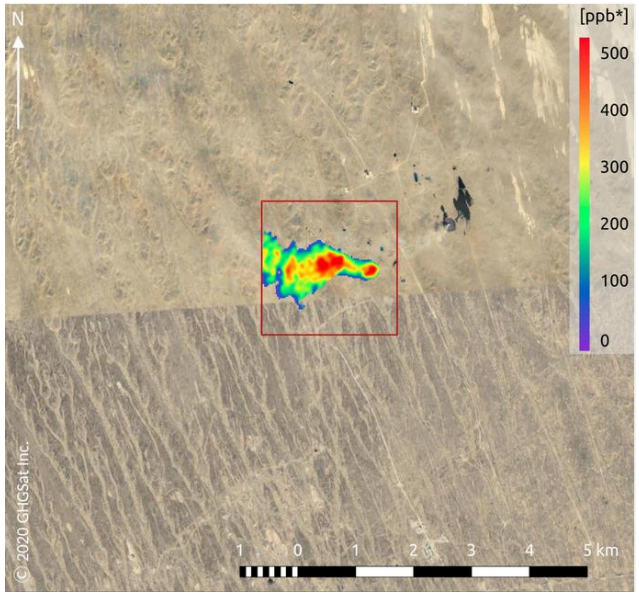


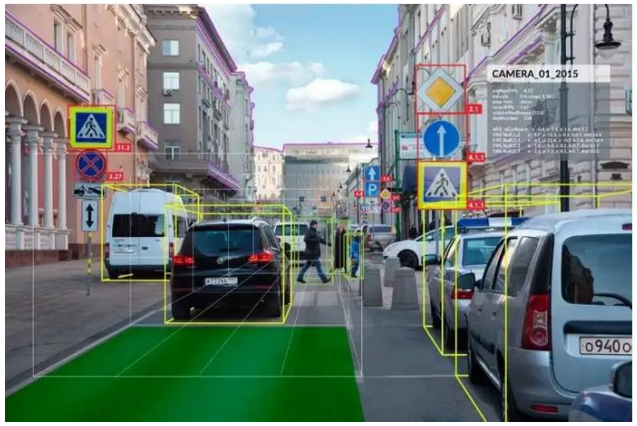
Image segmentation on X-Rays for clinical applications

https://www.frontiersin.org/articles/10.3389/fmedt.2021.767836/full



Satellite imagining for detection of gas leaks, detection

https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-5P/Monitoring_methane_emissions_from_gas_pipelines



Self-driving cars

https://becominghuman.ai/computer-vision-applications-in-self-driving-cars-610561e14118

2. Camera Control

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

2. Camera Control

3As (autofocus, auto-white balance, auto-exposure)

A DSLR gives the user a big range of parameters that provide the powerful art capacity of photography.

These parameters are set depending on:

- Semantic content of the scene: such as where do we want the observer to focus
- Properties of the scene: amount of light, distance to the target, etc

➢ But most smartphone users are not photographers!

➢ We can put ML to work to simulate common practices

2. Camera Control

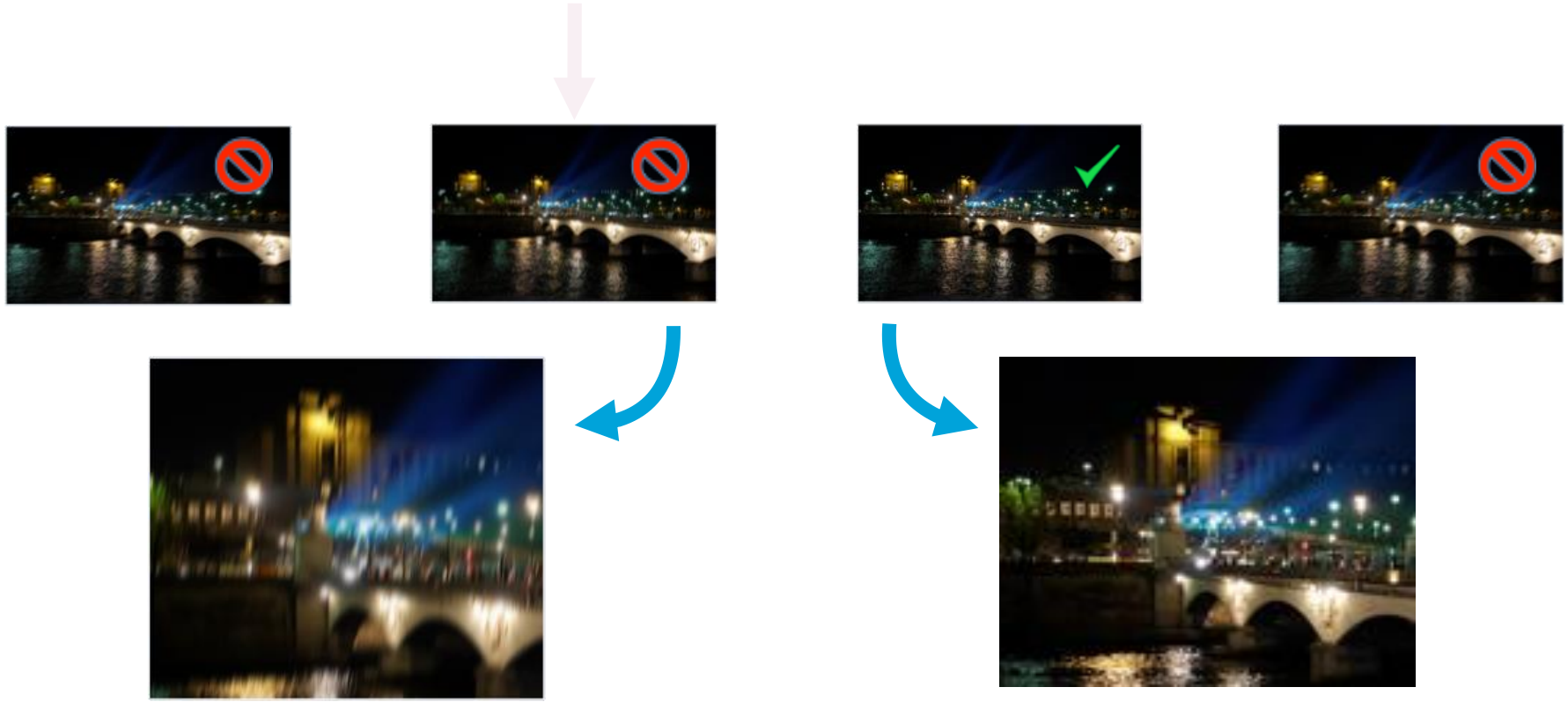
2. Camera Control: Subject detection

Our goal is to set up correctly the parameters of the camera so that the subject is well depicted:

He should be in-focus, correct exposure, etc

2. Camera Control

Image burst to automatically select the best moment



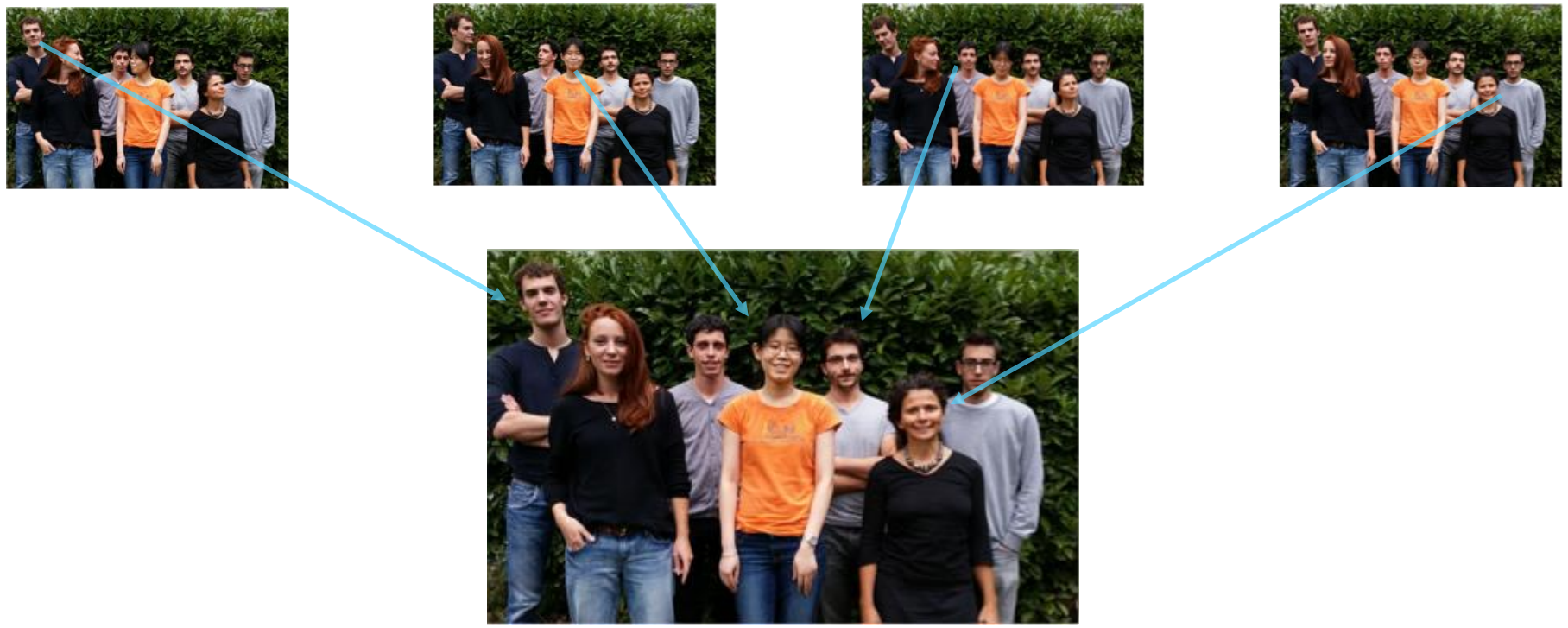
Select the image with less blur and more details

2. Camera Control: Face detection

Image burst with face detection to select the best moment

2. Camera Control: Image merge

Image burst with face detection to select the best moment



2. Camera Control: Semantic interpretation of the scene

<u>Identify the scene to adapt its treatement</u>

- The light level, the color of the light, the histogram are not enough to choose the parameters of the processing pipeline

→ We need to go further and interpret semantically the scene.

Convolutional Neural Network (CNN) trained to classify the scenes, objects

2. Camera Control: Semantic interpretation of the scene

AI for Bokeh

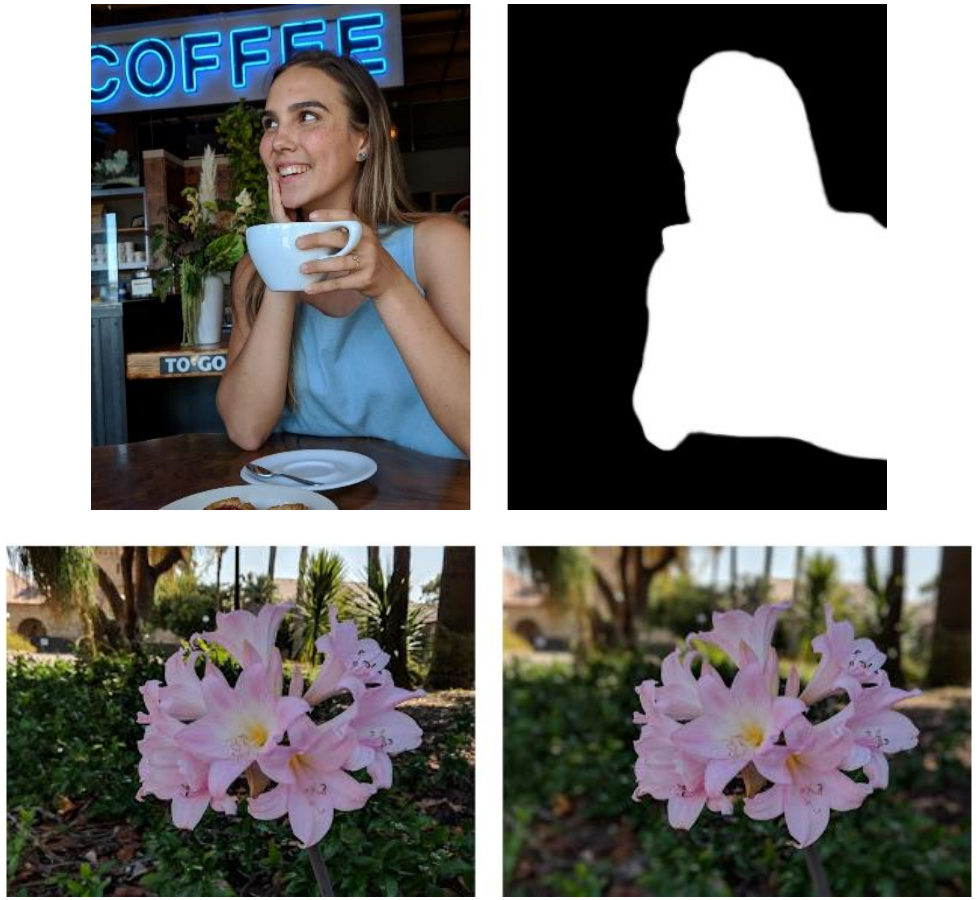
- Bokeh: A single-lens reflex (SLR) camera with a big lens has a shallow depth of field, meaning that objects at one distance from the camera are sharp, while objects in front of or behind that "in-focus plane" are blurry. Shallow depth of field is a good way to draw the viewer's attention to a subject, or to suppress a cluttered background

Images extracted from wikipedia

2. Camera Control: Semantic interpretation of the scene

AI for Bokeh

- How to simulate a Bokeh effect with a single camera (without the depth of the scene)?
- CNN trained to segment objects on the front from the background
- Just low-pass filter on the background
  - ➢ Can be effective, but very easy to spot



https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html

2. Camera Control: Conclusion

IA is not only improving the camera but converting it into a photographer.

3. Image Quality Assesment

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

Introduction - Image Quality

1. What is Image Quality?

   Perception of how a picture “looks good”, easily becomes a subjective matter

2. What influences Image Quality perception?

   - Shooting conditions: Camera performance depends on illuminant, illumination, motion in the scene, distance to the target, etc
   - Camera and its set up: optical parameters, sensor sensitivity, processing parameters…
   - Content of the scene
   - Knowledge of the person evaluating: IQ perception is different among image experts, professional photographers and consumers

Introduction - Subjective and objective evaluation

Subjective evaluation

- Real people give an opinion such as “big, small, heavy, light, cold, hot…”

Objective evaluation

- Usage of a measure that is neutral and operator independent: “2 cm, 15 meters, 150 kg, 2 g, -5°C, 50°C…”
- A device must provide figures (metrics) that are related to subject of measurement
- Normalizations may be necessary to have comparable metrics (e.g., cameras with different resolutions)
- Only addresses a set of metrics (i.e., some artifacts may be ignored)

Introduction - Subjective and objective evaluation

	Pros	Cons
Perceptual	• Related to perception	• Time consuming • Protocol is difficult to define • Costly • Difficult to compare with a previous batch
Objective	• Accurate • Cost effective • Repeatable • Comparable	• May miss artifacts; requires perceptual validation

We can use objective measurements as a “proxy” of perceptual ones to benefit from objective evaluation advantages

Introduction - Perceptual analysis

Evaluation of Image Quality attributes by a human operator

- Perceptual ≠ subjective
- **Methodology is key to get unbiased results**

Introduction - Challenges of perceptual analysis

- Neutrality

  Everyone should perform the same perceptual analysis on the same picture

- Relevancy

  The perceptual analysis must focus on aspects that are relevant for consumers and photographers

- Reliability and Repeatability

  A perceptual analysis protocol based on the capture of natural scenes which, by definition, are subject to variability in the shooting conditions (weather, light, time of day, …), must consider these factor

Introduction - Perceptual analysis methodology

IQA protocol

1. An analysis protocol
   What attributes/features to analyze and how to perform the analysis
2. A shooting protocol
   What scenes to shoot and how to shoot them

Provides

Neutrality: Any device can undergo the same testing and analysis procedure

Relevance : Results are meaningful for the end user

Reliability: Robust and repeatable results

Introduction – Why AI-based IQA models?

Even if with all the perceptual analysis methodology, on natural scenes, IQA perceptual evaluation

- is **a time-consuming** process
- can only be **relative** to images shot at the same moment as conditions are uncontrolled
- even when done by trained experts, a **variance** in results is possible

AI is a solution for having automated metrics that provide "perceptual like" results on complex areas and scenes.

3.1 What do we measure in IQA?

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

IQA: What are we measuring? What does it mean “looks good”?

1. **Aesthetic judgement**: “How far are we from photographic rules such as correct illumination, contrast, etc?”

2. **Visibility metrics**: “Can we see the difference between two images?” linked to Just Noticeable Differences (JNDs). Difficult to extend to wider transformations

3. **Impairment metrics**: Amount of distortion on an image due to noise, blur, compression, etc.

Goal: Create a psychophysical scale function

Psychophysical scaling refers to **the process of quantifying psychological events, especially sensations and perceptions**.

Goal

find the function: Physical stimuli → Scale of sensation, perception or any other mental event

Scaling by discrimination (developed by Fechner and Thurstone):
It is easier to detect differences between sensations but identify their magnitude is more difficult

MARKS, Lawrence E. et GESCHEIDER, George A. Psychophysical scaling. 2002.

Just Noticeable Difference: Weber law

The perception of differences between two stimuli depends on their range and value:

$$\frac{\Delta I}{I} = k, \tag{1}$$

where I is the original intensity of the particular stimulation, and ΔI is the addition to I required in order for this difference to be perceivable. Value k is a constant. The minimal ΔI is the Just Noticeable Difference (JND). This measure is statistical, that is to say, the JND is reported according to a group of people ("75% people noticed the difference"). JND is a measure of the limit of contrast perception, that is, evaluates if two signals that are slightly different will be perceived equally or not.

Just Noticeable Difference: Weber law

The perception of differences between two stimuli depends on their range and value:

“Can you see the difference between the background and the object?”

It is a discriminative experiment, thus the two evaluated images need to be very similar

Construction of JND perceptual scale

The JND scale is constructed statistically by encoding all the annotators answers.

- The probability of an image *difference* being qualified as visible is mapped to :

P(ΔI being visible) = 0.5 then the JND(ΔI)= 0.

P(ΔI being visible) = 0.75, JND(ΔI)=1

- The range of the final scale depends on the visual impact of the artifacts

Just Noticeable Difference does not capture Image Quality

The perception of differences between two stimuli depends on their range and value:

"Can you see the difference between the background and the object?"

It is a discriminative experiment, thus the two evaluated images need to be very similar

This is not what we are measuring in our experimental set up, and it is not what we are interested in!!

Experimental setup for image quality assessment

“Which image has more details (more fine visible details) that look natural? Artifacts should be penalized since they render the image unnatural. Some example of artifacts are ringing and over-sharpening, Moiré patterns - demosaicing artifacts- or over-smoothing in portraits -“waxy effect” on skin surfaces. “

- We are asking for a **preference judgement**
- There is an implicit reference to a **mental representation of the perfect image**
- We are not interested in knowing if there is a visible difference but whether this difference impacts our perception of quality
  - ➢ <u>Having more than 1 JND does **not** mean that the quality of our image has changed</u>

Short IQA experiment

Image Quality Assessment

*"Which image has **more details** (more fine visible details) that **look natural**? Artifacts should be penalized since they render the image unnatural. Some example of artifacts are ringing and over-sharpening, Moiré patterns -demosaicing artifacts- or over-smoothing in portraits -"waxy effect" on skin surfaces. "*

Which of the two images is better?

Image Quality Assessment

*"Which image has **more details** (more fine visible details) that **look natural**? Artifacts should be penalized since they render the image unnatural. Some example of artifacts are ringing and over-sharpening, Moiré patterns -demosaicing artifacts- or over-smoothing in portraits -"waxy effect" on skin surfaces. "*

Which of the two images is better?

Image Quality Assessment

Can you rank these images with respect to details on the highlights?

a)

b)

c)

d)

Image Quality Assessment

Can you rank these images with respect to details on the highlights?

a)

b)

c)

d)

Image Quality Assessment

Can you rank this image with respect to details on the highlights knowing the scale defined below?

0

5

10

Annotation methods highly impact final quality

Five methods ordered by difficulty of the task:

1. Pairwise with same content
2. Pairwise with different content
3. Ranking same content
4. Ranking with different content
5. Score assignment with same content

- We have a tradeoff between knowledge of the scene and quality estimation precision
- The more ambiguous the task, the higher the inconsistency of the annotations and the lower the precision of the final perceptual scale

3.2 Perceptual Image Quality Metric* Construction

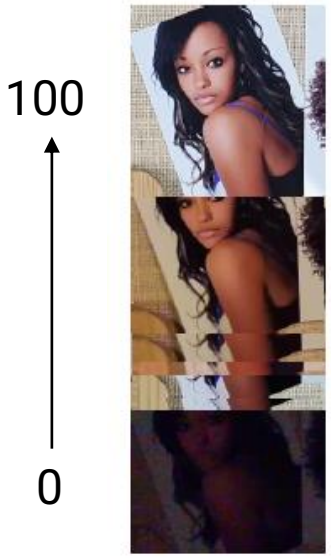
*Following definitions from: https://www.nist.gov/itl/ssd/software-quality-group/metrics-and-measures

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

Perceptual Image Quality metric construction: Mean Opinion Score

1. Definition of image quality anchors with associated scores



2. Annotator should assign a score value to each image taking into account the reference set

: 60

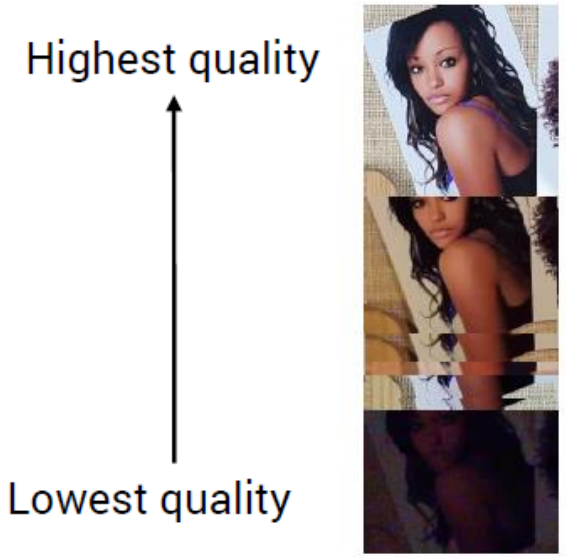
Image to assess

Mean opinion score (MOS): the score of the image is obtained by averaging all annotator's scores

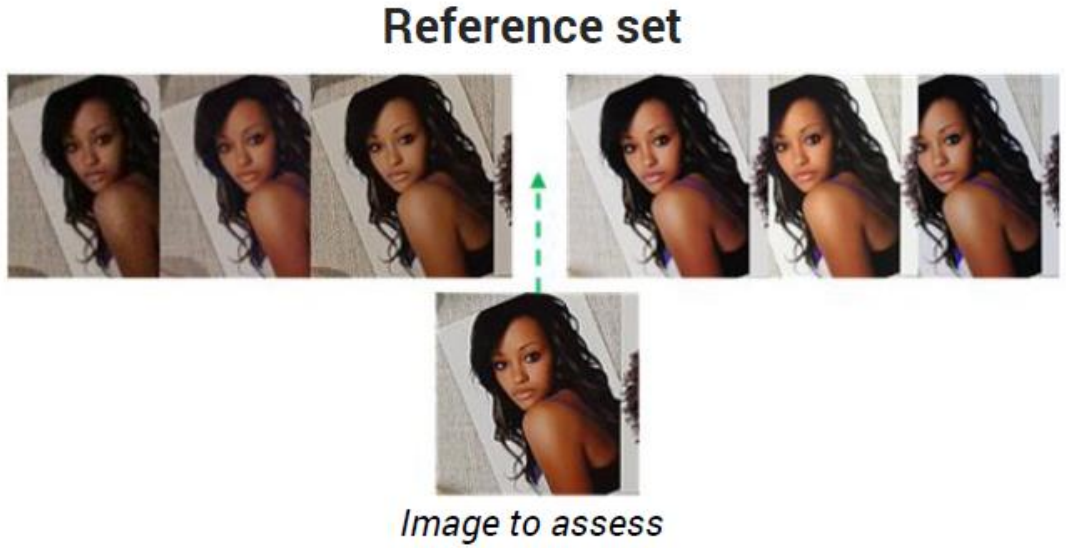
Drawbacks: Needs training to align all annotators with the predetermined scale, otherwise too wide inter annotator disagreement

Perceptual Image Quality metric construction: Ruler method

1. Definition of image quality anchors



2. Annotator should compare each image with the reference anchors and assign a position



The psychophysical scale is constructed with the ordering of the images averaging the position of all annotators

Drawbacks of the Ruler method

Pros:

- Once the anchors are selected, the annotations can be done fast
- Efficient for discriminating between 2 and 4 quality levels

Cons:

- Frontier effect around the anchors: Since the image quality is a function in high dimensions, it is difficult for the annotator to compare to just one series of images
- Lack of granularity in the final scale and dependent on the number and quality of anchor images selected

Pairwise comparison

The annotator is shown any pair of images in the database, and he/she needs to select the best one according to a given criteria (for instance blur worst than noise, fine grain noise better than coarse grain noise, etc)

- Intrinsically high dimensional (can take into account blur, over-sharpening, noise, etc)

M Perez-Ortiz, A Mikhailiuk, E Zerman, V Hulusic, G Valenzise, and R K Mantiuk. *From pairwise comparisons and rating to a unified quality scale*. IEEE Transactions on Image Processing, 29:1139–1151, 2019.

Pairwise comparison

The annotator is shown any pair of images in the database, and he/she needs to select the best one according to a given criteria (for instance blur worst than noise, fine grain noise better than coarse grain noise, etc)

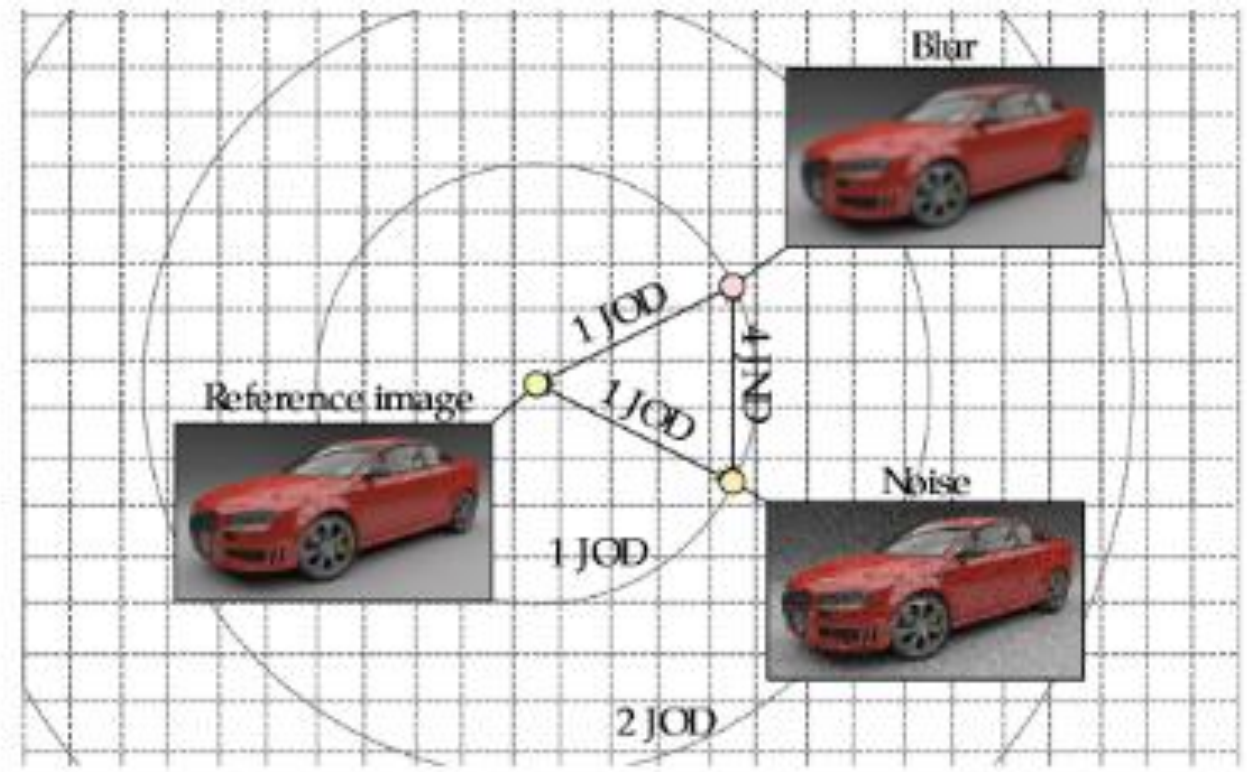
- Intrinsically high dimensional (can take into account blur, over-sharpening, noise, etc)

M Perez-Ortiz, A Mikhailiuk, E Zerman, V Hulusic, G Valenzise, and R K Mantiuk. *From pairwise comparisons and rating to a unified quality scale*. IEEE Transactions on Image Processing, 29:1139–1151, 2019.

Pairwise comparison for Just Objectionable Difference (JOD)

The annotator is asked to compare mentally with an ideal reference and evaluate if the images are “objectionable”

Just Objectionable Difference (JOD)



M Perez-Ortiz, A Mikhailiuk, E Zerman, V Hulusic, G Valenzise, and R K Mantiuk. *From pairwise comparisons and rating to a unified quality scale*. IEEE Transactions on Image Processing, 29:1139–1151, 2019.

Pairwise comparison for Just Objectionable Difference (JOD)

The JOD scale is constructed statistically by encoding all the annotators pairwise preferences.

- The probability of an image *A* being qualified as "of better quality" than image *B* is reflected in distances on the scale:

P(A>B) = 0.5 then the JOD(A,B)= 0.

P(A>B) = 0.75, JOD(A,B)=1

- The range of the final scale depends on the visual impact of the artifacts

M Perez-Ortiz, A Mikhailiuk, E Zerman, V Hulusic, G Valenzise, and R K Mantiuk. *From pairwise comparisons and rating to a unified quality scale*. IEEE Transactions on Image Processing, 29:1139–1151, 2019.

From pairwise comparison to psychometric scale

Estimated probabilities

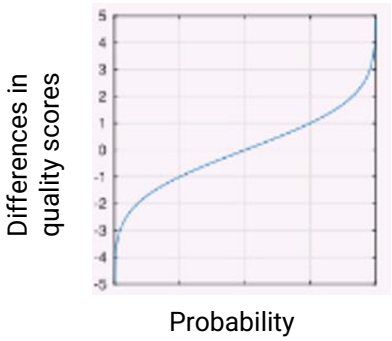
	A	B	C	D
A	0	1	0.97	1
B	0	0	0.23	0.03
C	0.03	0.77	0	0.33
D	0	0.97	0.66	0

Matrix of comparisons

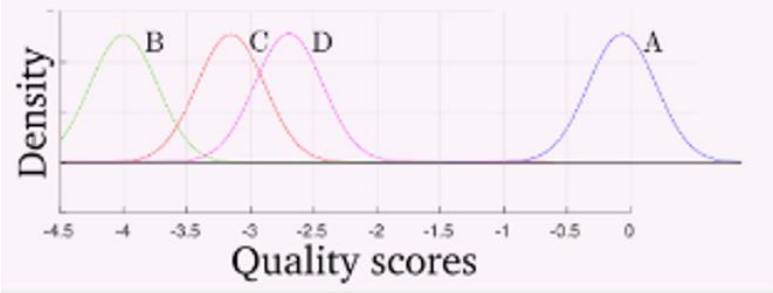
	A	B	C	D
A	-	30	29	30
B	0	-	7	1
C	1	23	-	10
D	0	29	20	-

30 annotators on PW experiment

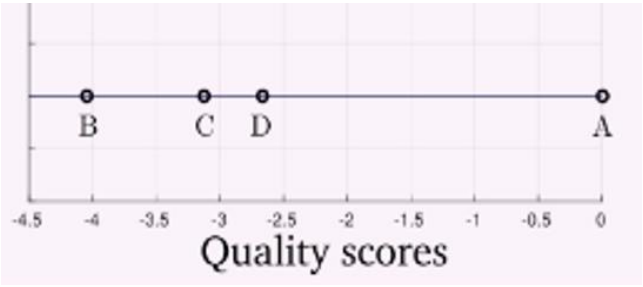
Estimated distances



Estimated distances



Scaling using MLE



M Perez-Ortiz, A Mikhailiuk, E Zerman, V Hulusic, G Valenzise, and R K Mantiuk. *From pairwise comparisons and rating to a unified quality scale*. IEEE Transactions on Image Processing, 29:1139–1151, 2019.

3.3 Generic IA-based IQA models

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

Generic AI for image quality assessment

Common numerical full reference measures of Image quality

- Mean Square Error:

$$MSE(x,y) = \frac{1}{WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}\|x(i,j)-y(i,j)\|^2$$

- Peak-signal to noise ratio:

$$PSNR(x,y) = 20\log_{10}\left(\frac{255}{MSE(x,y)}\right)$$

- Structural similarity Index: Perception-based, it assumes degradations are perceived changes in structural information

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

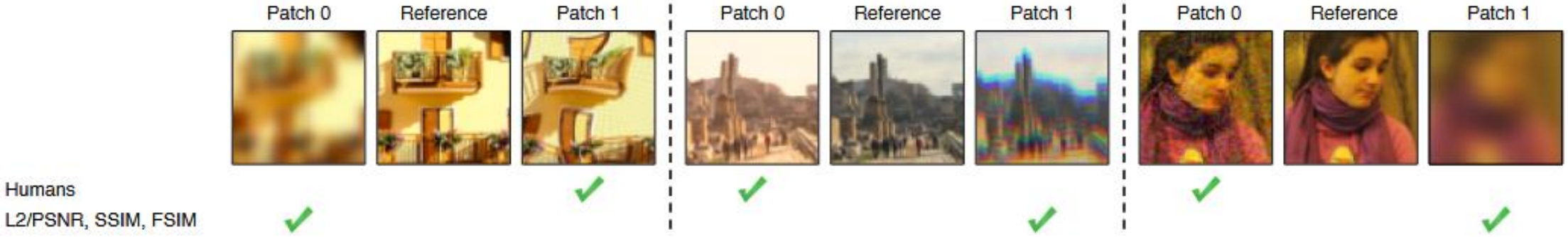
μ_x μ_y : mean values, related to luminance

σ_x σ_y σ_{xy} : std, related to contrast

Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." *IEEE transactions on image processing* 13.4 (2004): 600-612.

Generic AI for image quality assessment

But these measures do not capture the human image quality perception
(a.k.a do not correlate with our psychophysical JOD scale)



Which patch (left or right) is “closer” to the middle patch in these examples?

Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Generic AI for image quality assessment

Learned Perceptual Image Patch Similarity (LPIPS)

Idea: Neural features model low-level perceptual similarity very well and can be used to full-reference quality assessment

1. Use the neural representation of the two input images trained for other tasks (for instance classification)
2. Learn the mapping between the neural distance and the perceptual estimation

Parameters to be learned

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw})||_2^2$$

Neural representation of image (i.e. AlexNet or VGG)

Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Generic AI for image quality assessment

Learned Perceptual Image Patch Similarity (LPIPS)

Idea: Neural features model low-level perceptual similarity very well and can be used to full-reference quality assessment

1. Use the neural representation of the two input images trained for other tasks (for instance classification)
2. Learn the mapping between the neural distance and the perceptual estimation

Training data set :

- Reference images
- Artificial distortions: colorization, video deblurring, superresolution, frame interpolation, neural distortions
- The scores measure the level of perceived distortion between images
- Patches of 64x64

Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

Generic AI for image quality assessment

Learned Perceptual Image Patch Similarity (LPIPS)

Idea: Neural features model low-level perceptual similarity very well and can be used to full-reference quality assessment

Drawbacks for Image Quality assessment:

- Data set trained on artificial distortions: representative of real camera outputs?
- Perceptual similarity on distortions, in general, is not quality assessment
- Patches of 64x64 does not help when evaluating a complete image

Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

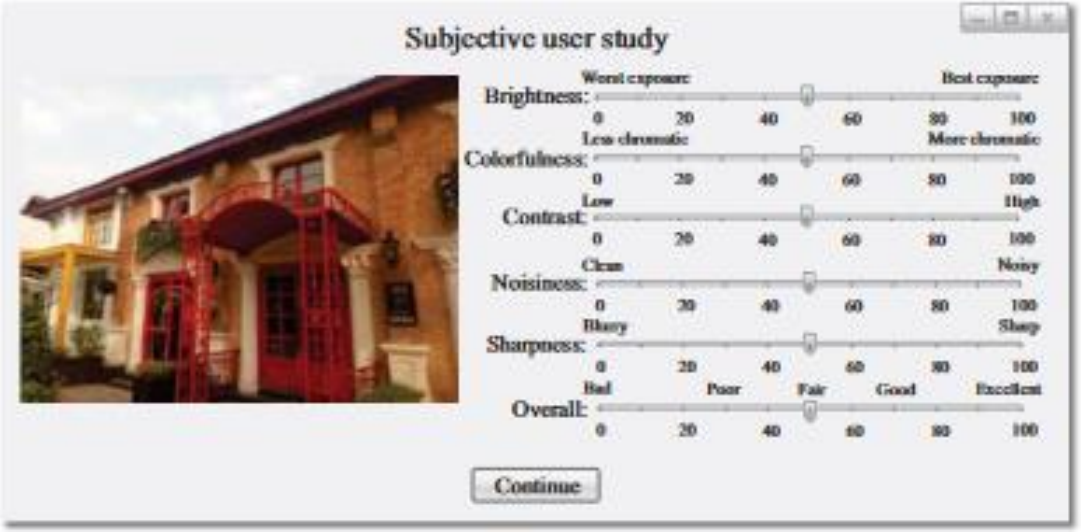
Generic AI for image quality assessment

Perceptual Quality Assessment of Smartphone Photography (SPAQ)

Blind (no reference) image quality assessment

Data set :

- Smartphone cameras:
  - 1k smartphones HR images,
  - 10k realistic distorted images: over-exposure, motion blur, contrast reduction, etc
- MOS annotations
- Image attributes: brightness, colorfulness, contrast, noisiness, sharpness, overall quality



Fang, Yuming, et al. "Perceptual quality assessment of smartphone photography." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Generic AI for image quality assessment

Perceptual Quality Assessment of Smartphone Photography (SPAQ)

Blind (no reference) image quality assessment

They propose two models:

- Baseline Model :
  - ResNet-50 adapted for regression with a scalar output and trained with a L1-loss function
- Multi-task learning: MT-A
  - Same structure but trained on all features at the time (6 outputs)
- Multi-task learning: MT-E
  - Taking advantage of the information present in the EXIF files, they propose to also use it in the prediction (focal length, f-number, ISO)

Fang, Yuming, et al. "Perceptual quality assessment of smartphone photography." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Generic AI for image quality assessment

Perceptual Quality Assessment of Smartphone Photography (SPAQ)

Blind (no reference) image quality assessment

Results Spearman's Rank Order Correlation coefficient (SRCC):

Attribute	Image attribute scores from humans	Image attribute scores by MT-A
Brightness	0.784	0.704
Colorfulness	0.844	0.760
Contrast	0.874	0.786
Noisiness	0.893	0.832
Sharpness	0.958	0.904

Table 2. SRCC results between MOSs and image attribute scores from humans and MT-A (our proposed computational model), respectively.

Model	BL	MT-E	MT-A	MT-S
SRCC	0.908	0.926	0.916	0.917

Fang, Yuming, et al. "Perceptual quality assessment of smartphone photography." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Generic AI for image quality assessment

Perceptual Quality Assessment of Smartphone Photography (SPAQ)

Blind (no reference) image quality assessment

Drawbacks:

- Not evaluating the same content: Difficulty in the image assessment annotation task means strong inter-annotator disagreement
- MOS annotations lack precision

Fang, Yuming, et al. "Perceptual quality assessment of smartphone photography." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

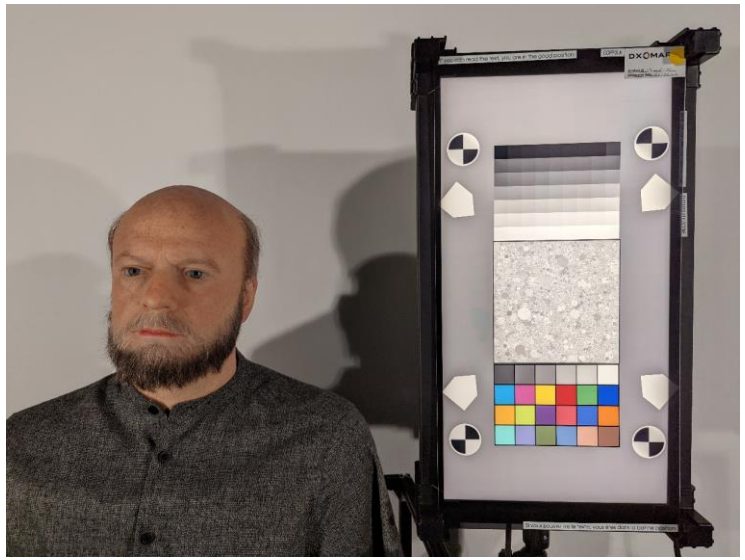
3.4 Precise AI based IQA models

Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

Precise AI based IQA models : AI texture-noise evaluation

Doing AI measurements on complex laboratory scenes allows to be closer to natural scene content while keeping controlled conditions for repeatable shooting, and thus comparable non-relative results.



Realistic mannequin with HDR conditions

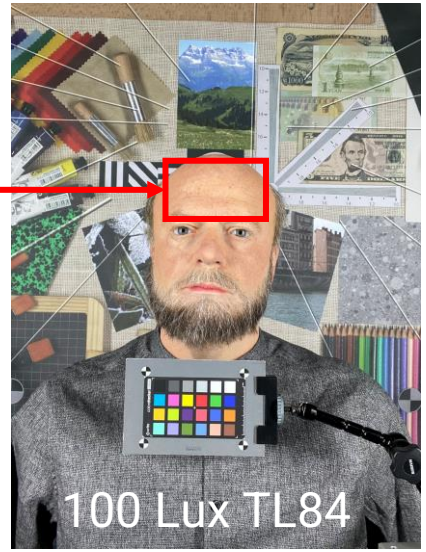
Complex charts with everyday objects

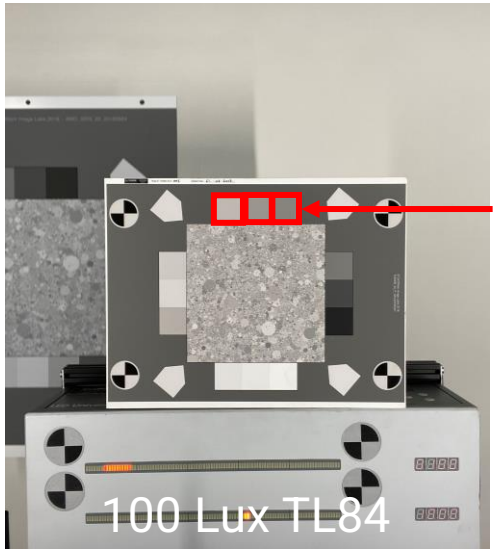
Different content, different behavior: test with noise rendering

- Does the presence of a face really trigger a different noise rendering?

→ For every device, take pictures of a deadleaves chart in the exact same conditions as the realistic mannequin.

Noisiness Annotation (JOD scale)
The lower, the more noise

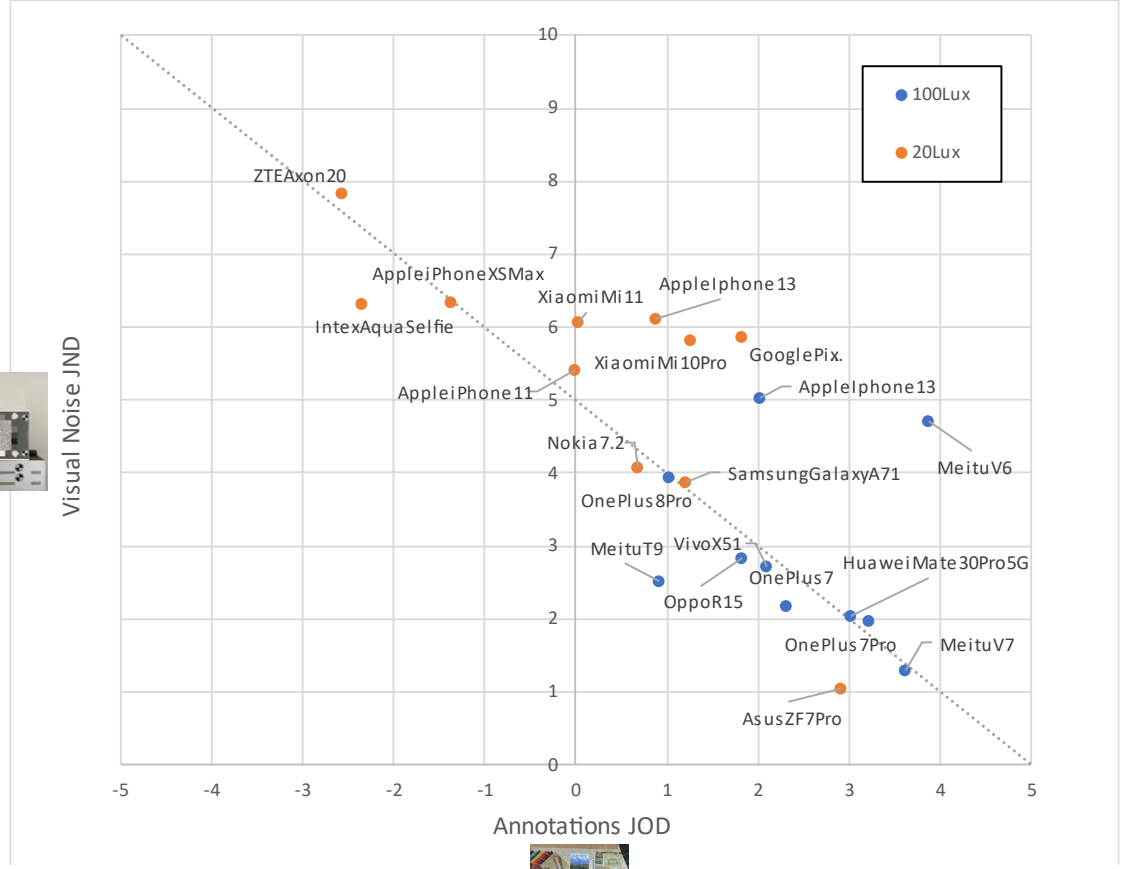




Visual Noise (JND scale)
The lower, the less noise

Different content, different behaviour: test with noise rendering

Noise perceptual annotation on Realistic Mannequin
vs measure of visual noise on Deadleaves

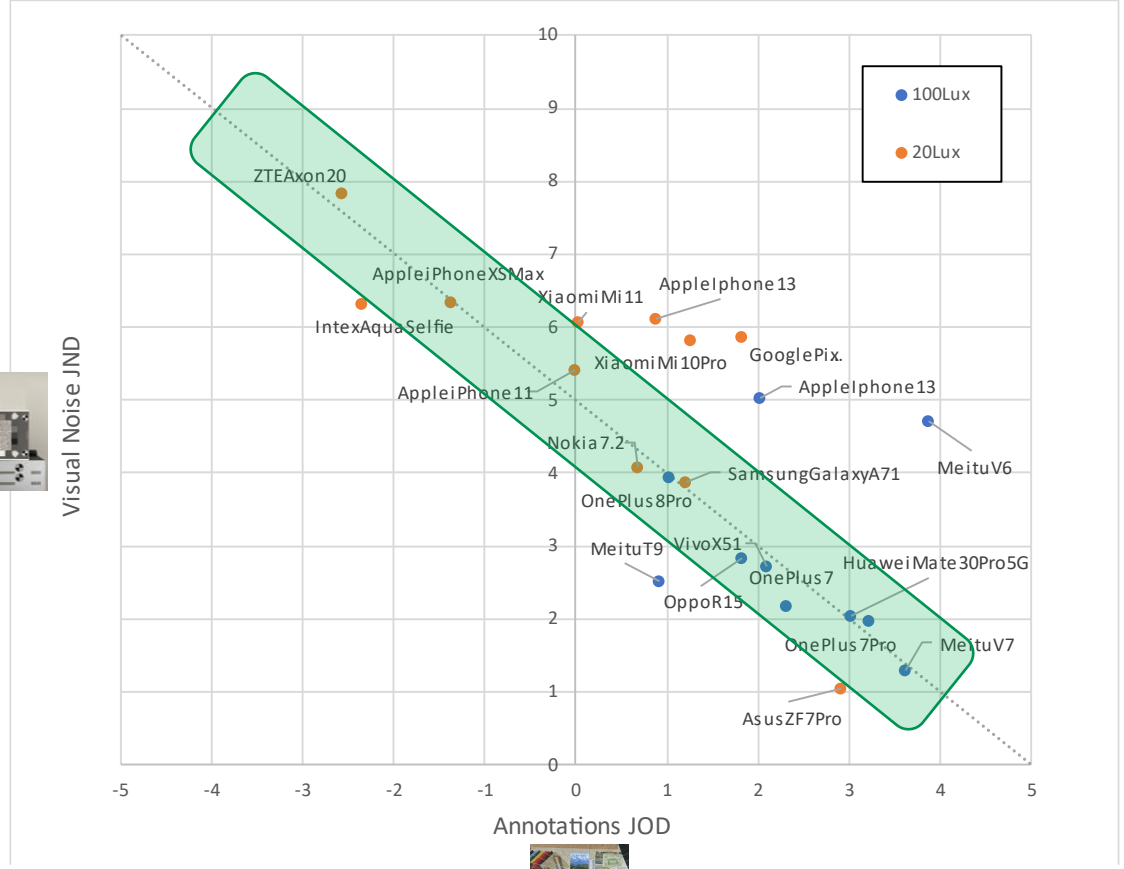


For every device, we compare to Visual noise on a crop on the Deadleaves chart

- lightness L* of the deadleaves crop is the closest possible to the lightness L* of the crop of the forehead of the mannequin
- **Visual Noise is mapped to a JND of noisiness** scale (the lower the value the lower the perceived noise).
- **IQA on noise rendition on Realistic Mannequins** images are annotated in quality units (the lower the value the greater the perceived noise).
  - Viewing conditions during the annotations are the same (image viewed with a cutoff frequency of 30 cycles per degree)

Different content, different behaviour: test with noise rendering

Noise perceptual annotation on Realistic Mannequin
vs measure of visual noise on Deadleaves

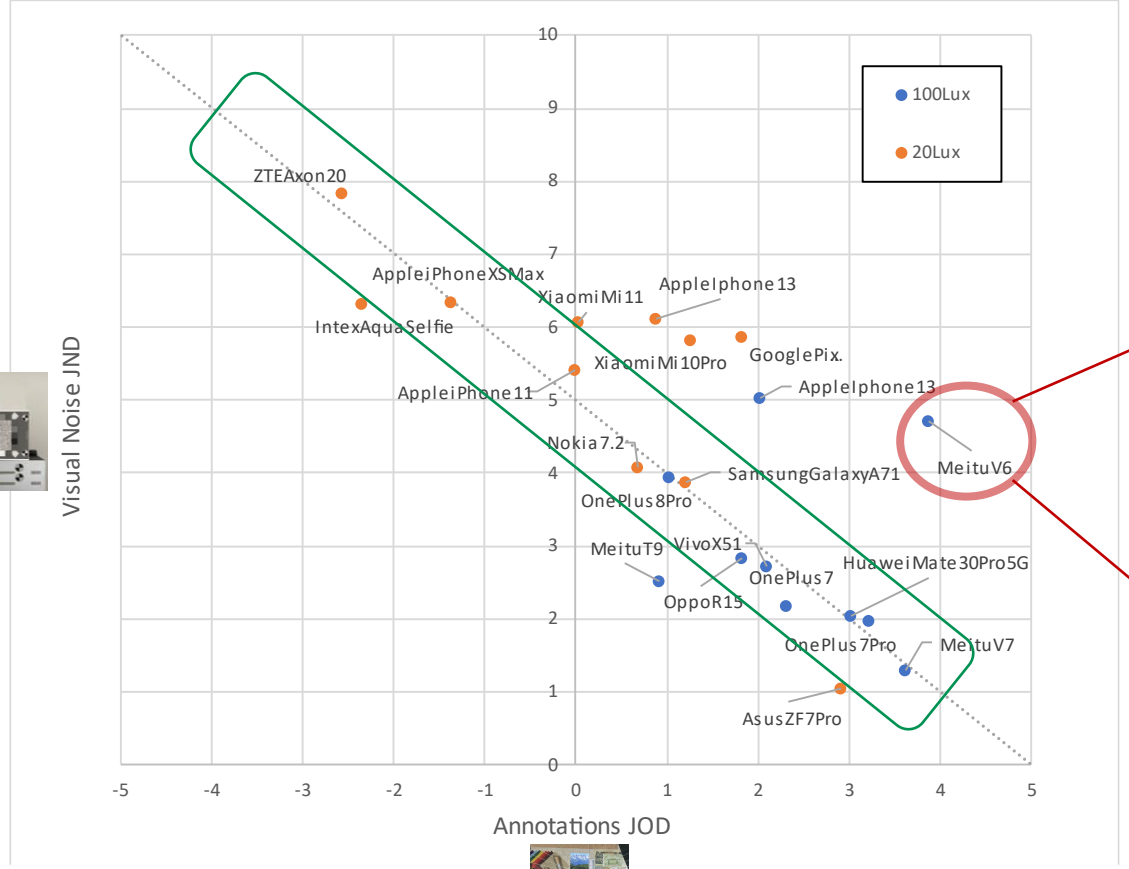


Most devices are aligned:

The relative perception of noise on the forehead of the mannequin and a uniform patch is generally similar

Different content, different behaviour: test with noise rendering

Noise perceptual annotation on Realistic Mannequin
vs measure of visual noise on Deadleaves

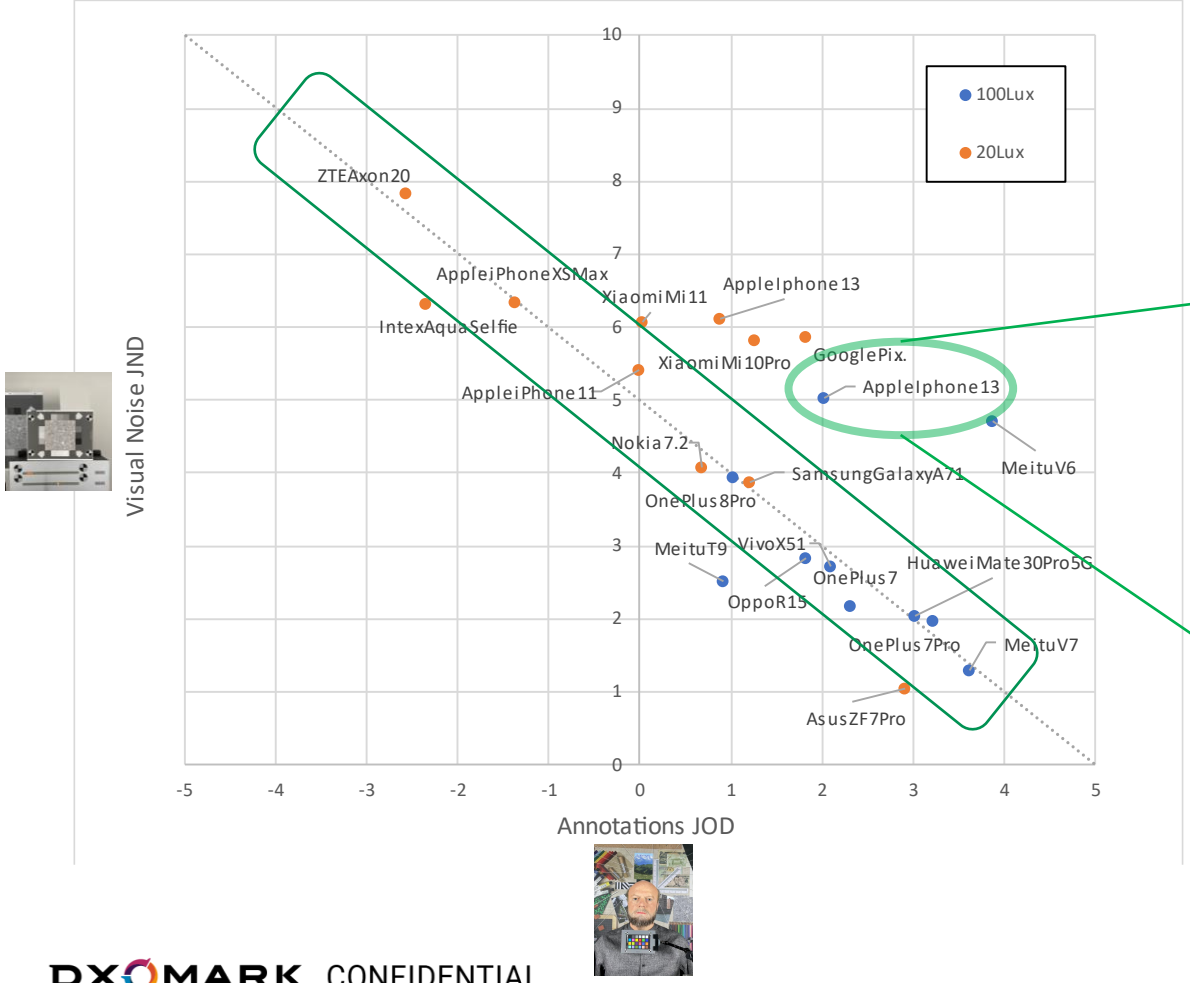


Not always the case

- Skin strongly denoised (wax effect)
- Noise still present in the flat patches

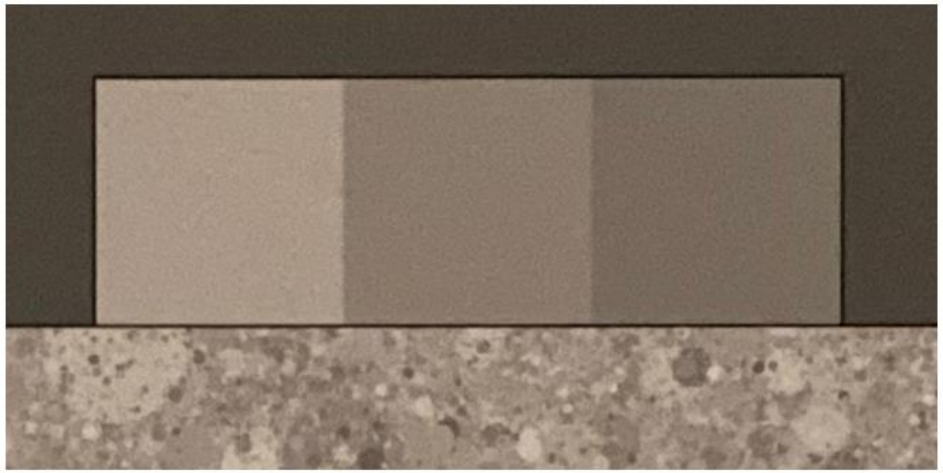
Different content, different behaviour: test with noise rendering

Noise perceptual annotation on Realistic Mannequin vs measure of visual noise on Deadleaves



Not always the case

- Perception of noise on textured surfaces is different from flat surfaces: Slight noise on the mannequin was judged as less objectionable



Different content, different behaviour: test with noise rendering

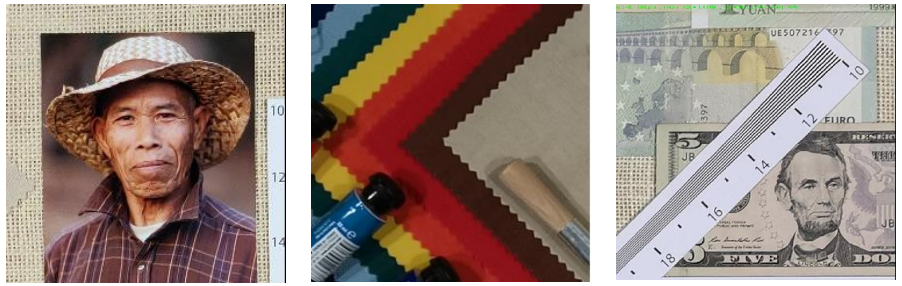
- Cameras treat differently images with different content: faces vs synthetic content
- Our perception of certain features also depend on the content: slight noise is not perceived equally on a flat region that on a face

In order to thoroughly test devices we need diverse set ups

3.4 Precise AI based IQA models : Laboratory

DXOMARK Chart - Dataset

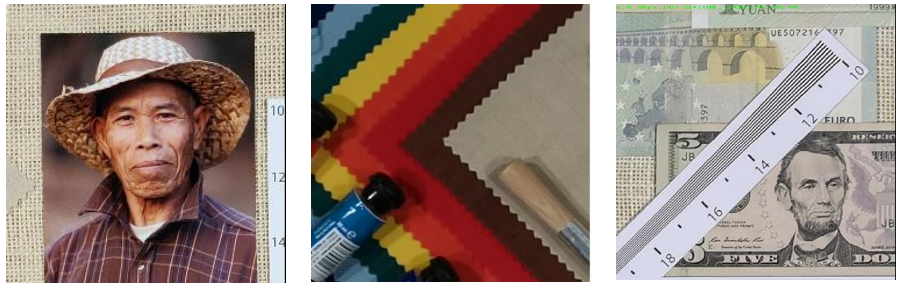
Analysis for Texture



Analysis for Noise

DXOMARK Chart - Dataset

Analysis for Texture



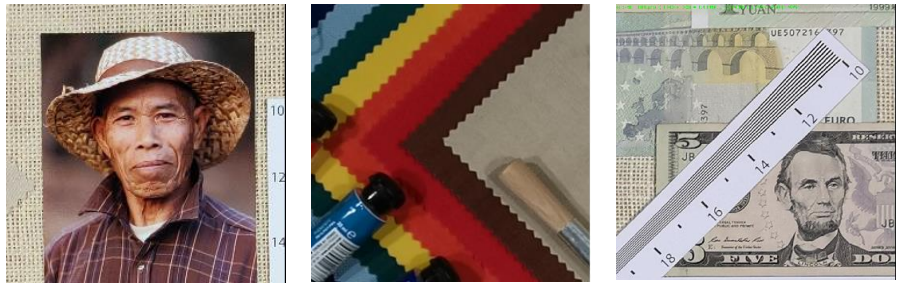
Analysis for Noise

Different Lighting conditions

5 Lux A | 20 Lux A | 100 Lux A | 300 Lux TL84 | 1000 Lux D65

DXOMARK Chart – Texture Model

Analysis for Texture

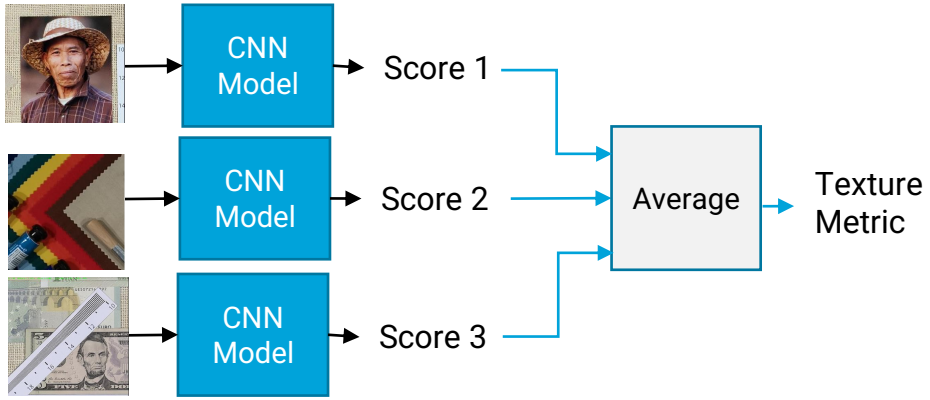


Data set:

- Selection of the DXOMARK image database:
  - Images of different qualities from 140 devices
  - 5 lighting conditions
  - No-brand overlap split between train and test
  - Images are registered
- Annotations from image quality experts

Model:

- ResNet-50 trained to regress the score quality
- Final result is the average score of all patches



	SROCC (Spearman Rank Correlation)	KROCC (Kendall Rank correlation)
model	0.9	0.728

DXOMARK Chart – Noise Model

Data set:

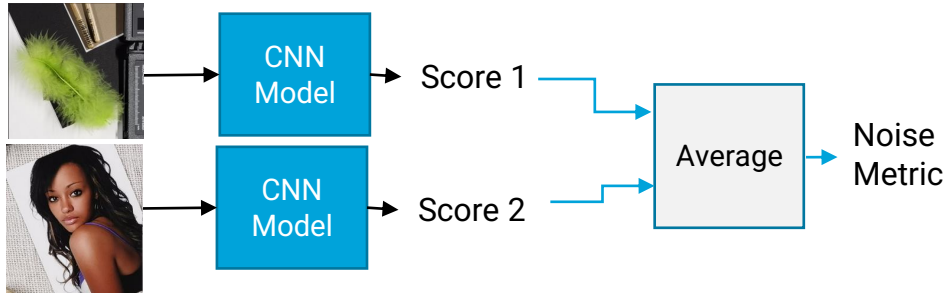
- Selection of the DXOMARK image database:
  - Images of different qualities from 293 smartphones
  - 6 distinct manufacturers
  - No-brand overlap split between train and test
  - Images are registered
- Annotations from image quality experts

Model:

- ResNet-50 trained to regress the score quality
- Final result is the average score of all patches

	SROCC (Spearman Rank Correlation)	KROCC (Kendall Rank correlation)
model	0.904	0.734

Analysis for Noise



3.4 Precise AI based IQA models : Natural Scenes

AI texture evaluation on natural scenes

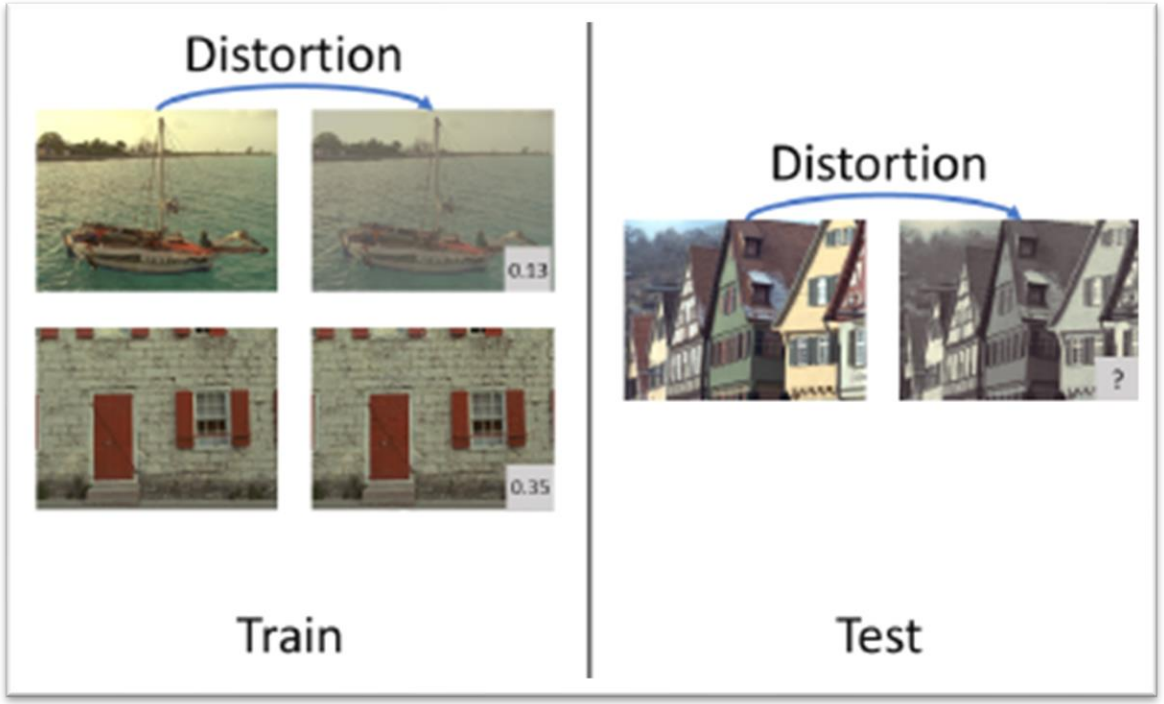
We know smartphone cameras adapt to the scene content, thus we need to evaluate on natural content

- Define a shooting protocol:
  1. Select a predefined set of scenes
  2. Select regions of interest that are invariant along all the images in the scene (not dependent of the weather, season, etc)

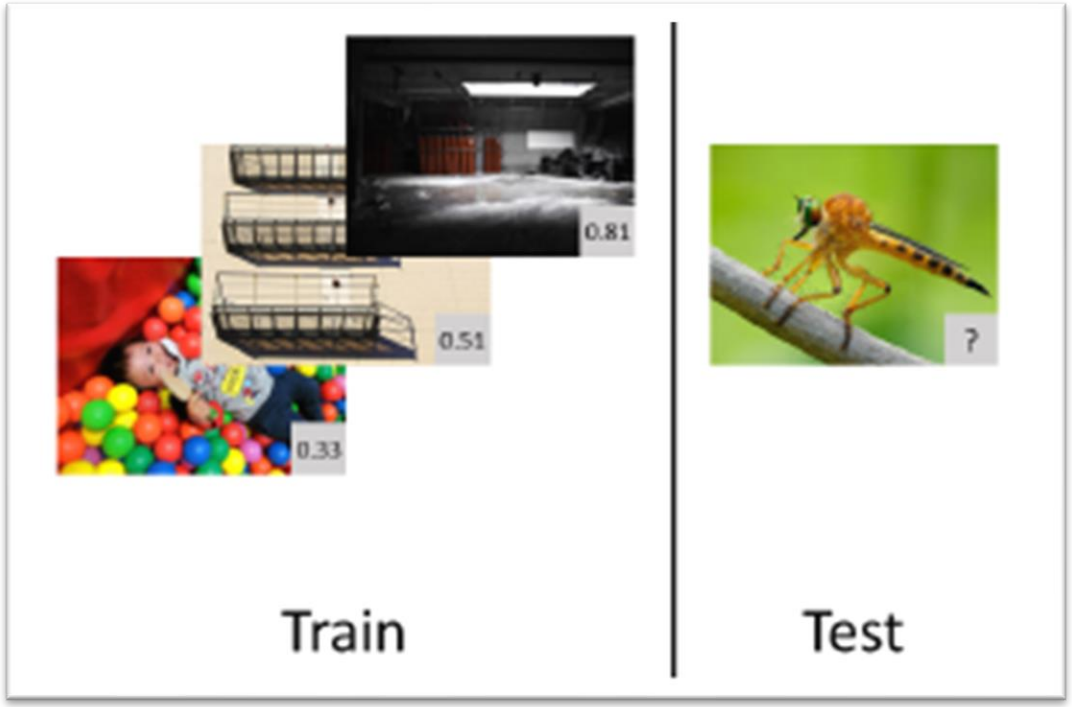
- Define a testing protocol:
  1. Data set creation with a wide quality variety of devices
  2. For every scene, select a reference device (DSLR, for instance)
  3. Split scenes into train and test to measure the generalization capacity of the model

AI texture evaluation on natural scenes: Reference-based IQA

Full reference IQA



No reference IQA

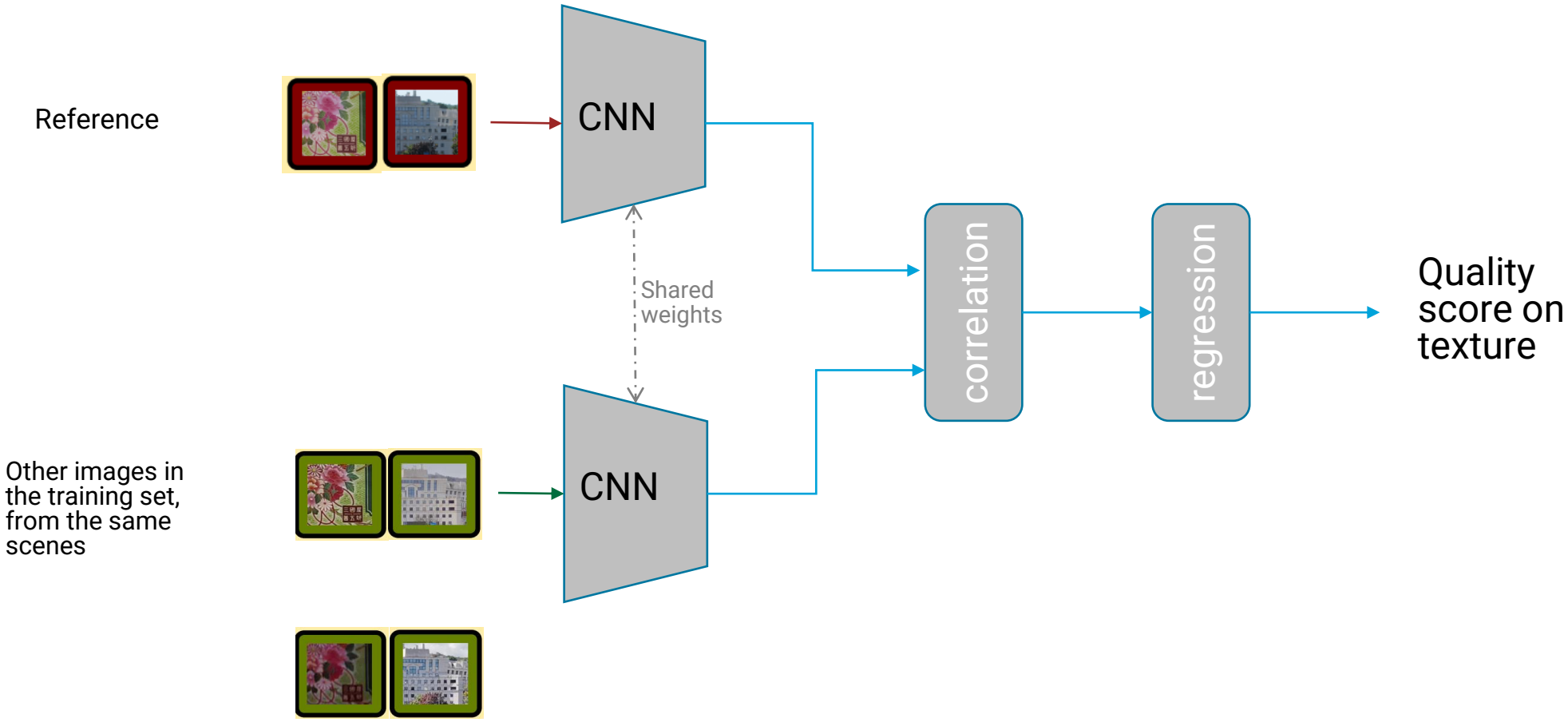


AI texture evaluation on natural scenes: Reference-based IQA

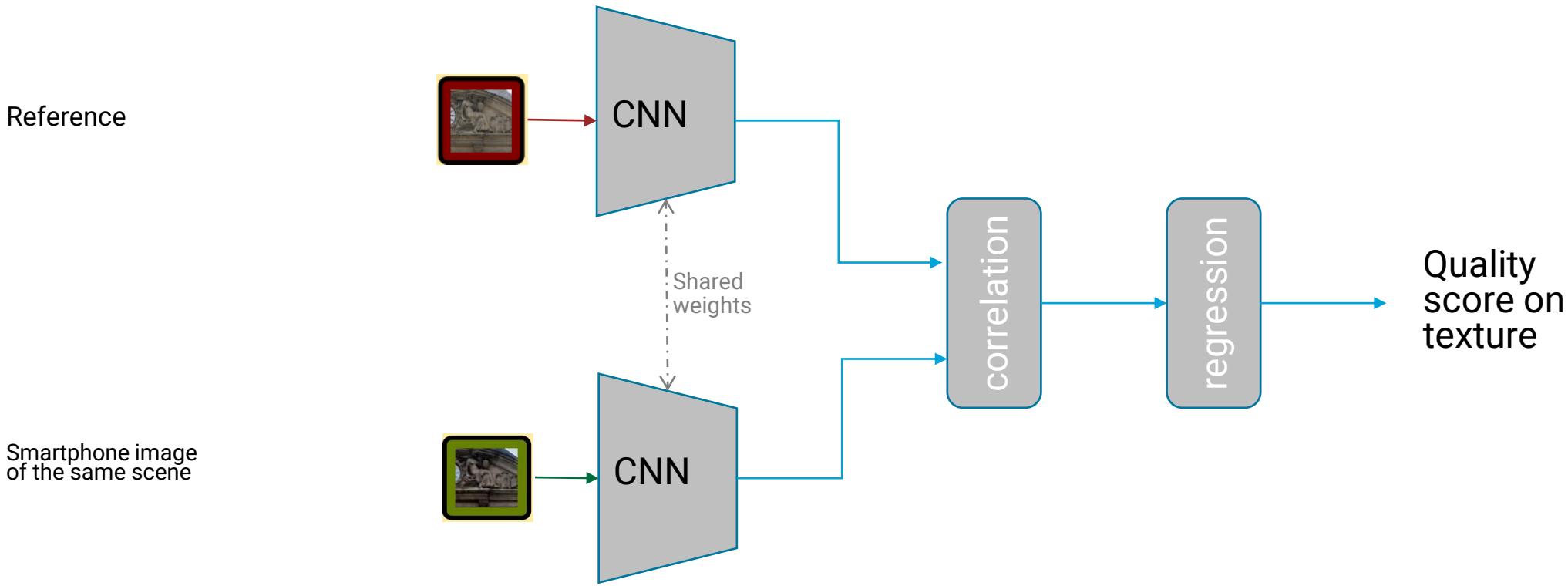
Dataset creation:

1. Find Region-of-interest for reference and target image
2. Align images with reference registration and upscaling
3. Annotation: Pairwise comparison and JOD generation

AI texture evaluation on natural scenes: Training time



AI texture evaluation on natural scenes: Inference



AI texture evaluation on natural scenes: Results

Training and evaluating

- No semantic overlap between train and tests set: the split is based on the scenes
- Training set : to all scenes but one (used as a test set)
- This scenario is particularly useful to the generalization ability of the different models

Backbone	SROCC	KROCC
ResNet-18	0.898 ± 0.003	0.755 ± 0.005
ResNet-50	0.892 ± 0.004	0.744 ± 0.006
MobileNet-V2	0.893 ± 0.004	0.746 ± 0.007
VGG-16	$\mathbf{0.900} \pm 0.006$	$\mathbf{0.760} \pm 0.009$

3.5 AI based portrait quality assessment (PQA)

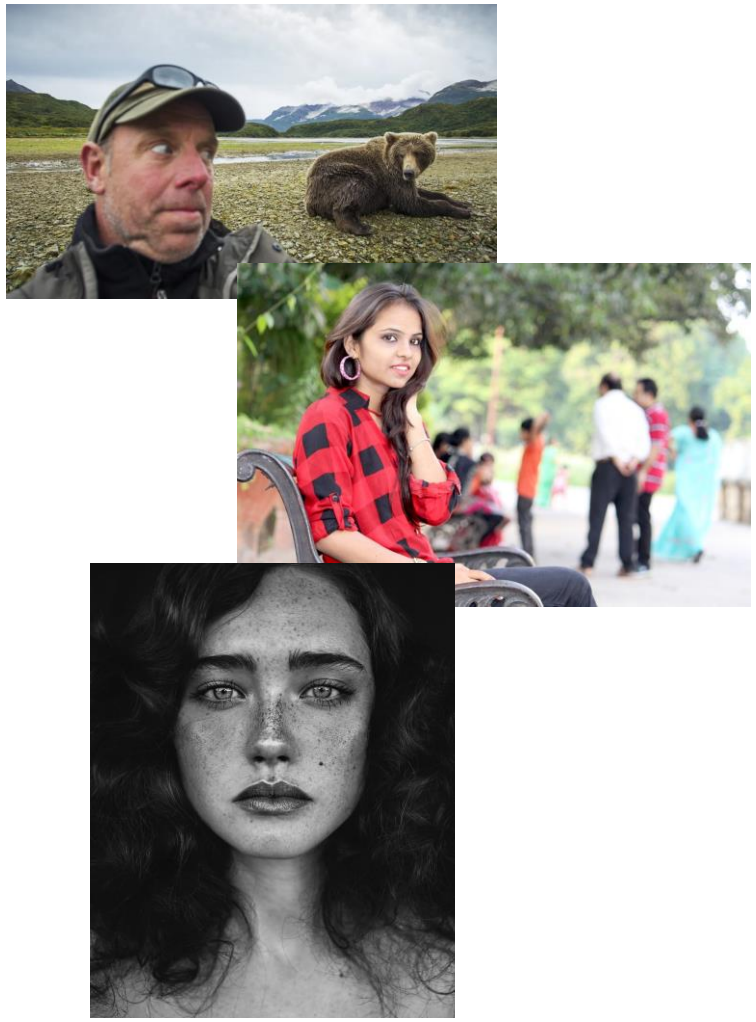
Artificial Intelligence for smartphones and image quality

1. AI and Deep Learning
2. Camera Control
3. Image Quality Assesment
    1. What do we measure in IQA?
    2. Perceptual Image Quality Metric Construction
    3. Generic IA-based IQA models
    4. Precise AI based IQA models
        1. Laboratory
        2. Natural scenes
    5. AI based portrait quality assessment (PQA)
        1. Laboratory set up
        2. Portraits in the wild

PQA – Portrait use case

- Portrait photography is a widely common use case:
  - Social media
  - Advertisements
  - Official documents

- Portraits are known to be challenging for image enhancement algorithms:
  - Various characteristics:
    - Skin tone, Gender, Age
    - Facial hair, Face shape, Facial features, Facial expressions
    - Clothing, backgrounds, etc.
  - No universal agreement on good portrait representations:
    - Skin tone, skin smoothness and face exposure can depend on the culture
    - Compromising the quality of the background in contrast to the subject is relative to the use case, etc.



PQA – Portrait use case

- Multiple factors can affect the quality of portrait images
  - Environment: Lighting conditions, Scene Dynamics, Scene content, etc.
  - Subject: Framing, Distance from the camera
  - Camera: Focal length, Sensor size, Settings, etc.

- Smartphone cameras compensate these difficulties mainly with software and AI:
  - This leads to weird unrepresented local and global distortions and artefacts
  - Image enhancement and correction is thus local and content dependent

PQA – Portrait use case

- The behavior of smartphone cameras is content dependent:
  - Same shooting conditions, different outcomes

- On portraits, smartphones tend to smoothen the skin, over sharpen texture and sometimes overexpose the face.

- Skin strongly denoised (wax effect)
- Noise still present in the flat patches

PQA – Laboratory setup

- What’s a better subject to evaluate a camera’s portrait performance than a real person?

PQA – Laboratory setup

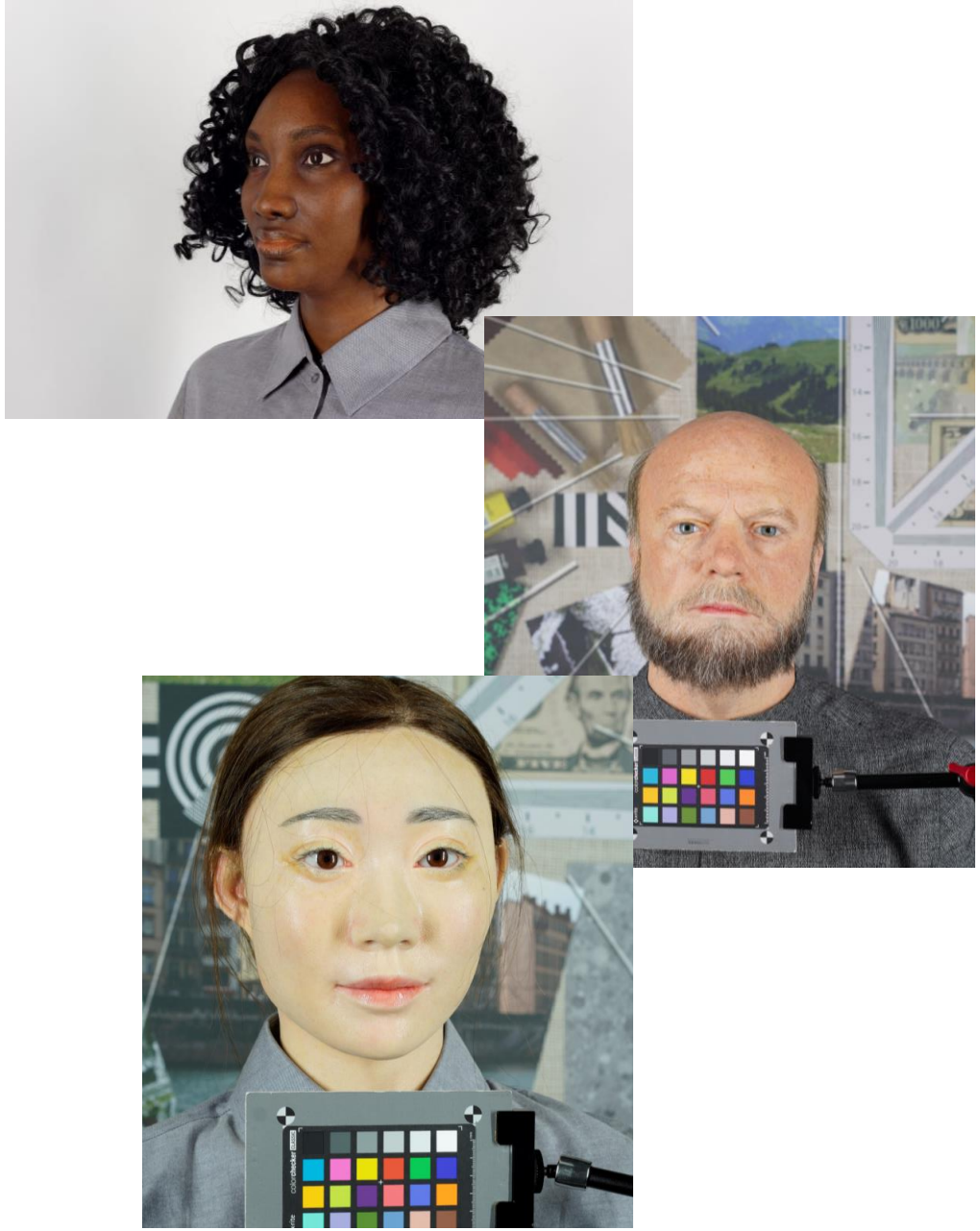
- What’s a better subject to evaluate a camera’s portrait performance than a real person?

‘A realistic mannequin!’
- DXOMARK -

PQA – Laboratory setup

- What’s a better subject to evaluate a camera’s portrait performance than a real person?

‘A realistic mannequin!’
- DXOMARK -



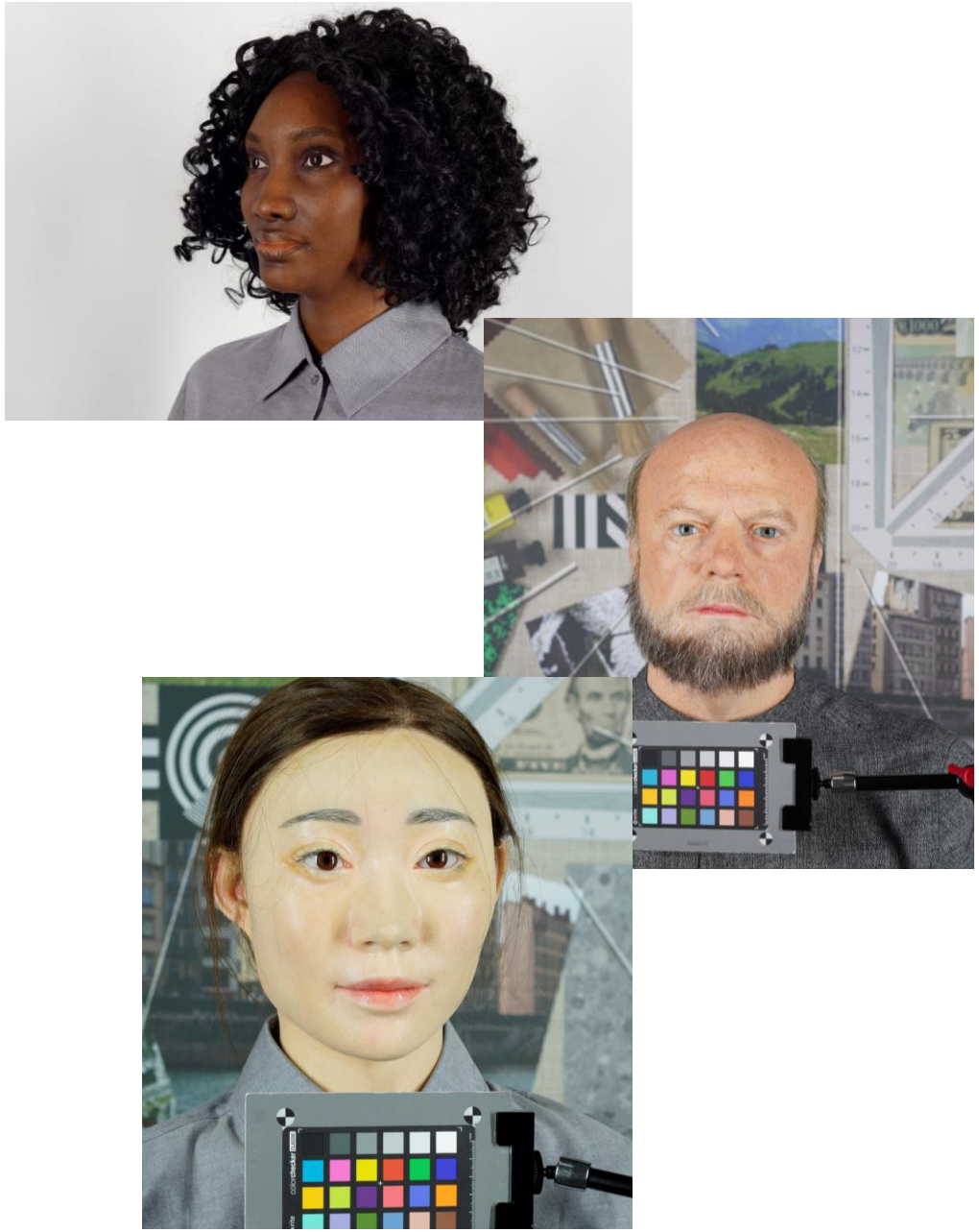
PQA – Laboratory setup

- What's a better subject to evaluate a camera's portrait performance than a real person?

'A realistic mannequin!'
- DXOMARK -

- Why?
  - Fixed content means repeatable measurement
  - Simulate a real portrait scenario
  - Captures complex behaviour of cameras (triggers specific pipelines for portraits)

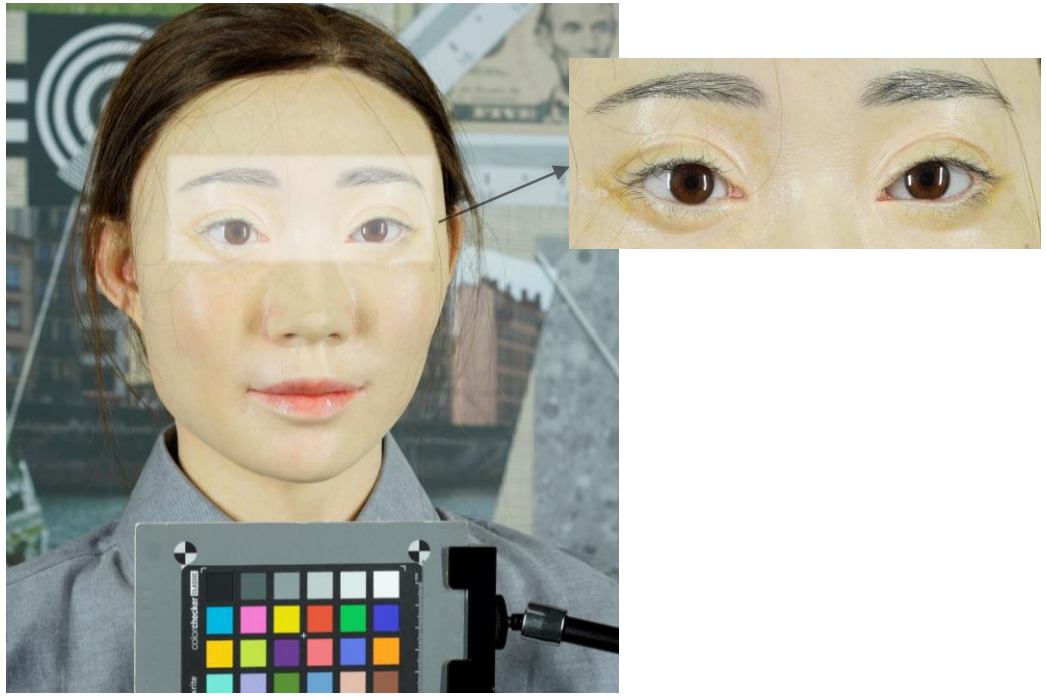
- How?
  - Dataset of RM images, shot with different devices, lighting conditions and shooting distances
  - Subjective image quality scores based on pairwise comparisons
  - Convolutional neural networks



PQA – Detail preservation on realistic mannequins [1]

- Predefined regions of interest:
  - Fixed aspect ratio
  - Upscaled to the highest resolution
  - No resize distortions

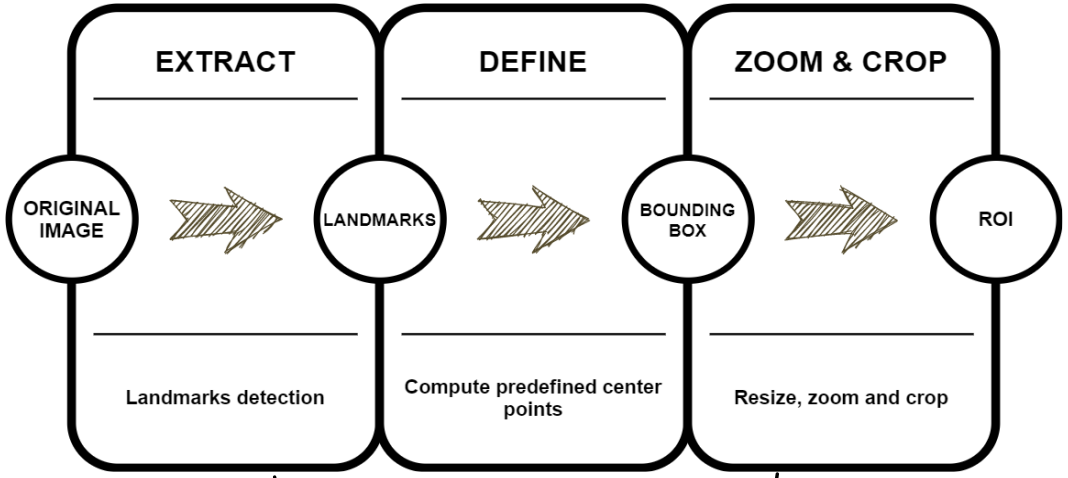
- Data annotations:
  - Evaluation of fine detail preservation
  - Pairwise comparisons mixed with active sampling based on information gain [2]:
    - Maximize Kullback-Leibler divergence between the current and the resulting distribution after every possible comparison
    - Final scores are on a Just Objectionable difference (JOD) scale (Thurstone case V)
  - Fixed viewing condition, aligned with the human eye perception (Vcut = 30cpd)

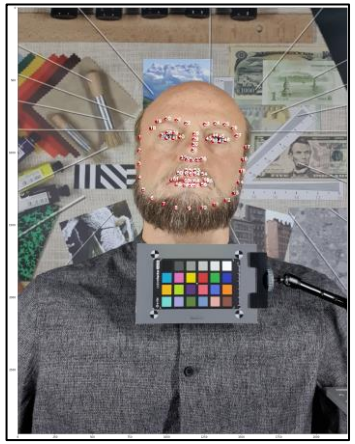


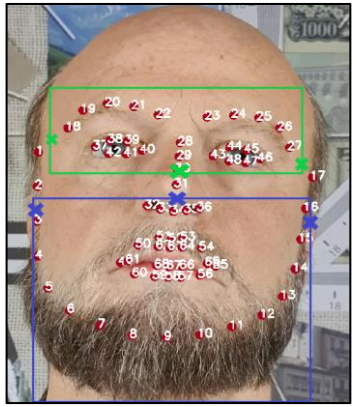
[1] Nicolas, Chahine, and Belkarfa Salim. "Portrait Quality Assessment using Multi-Scale CNN." London Imaging Meeting. Vol. 2. Society for Imaging Science and Technology, 2021.
[2] Mikhailiuk, Aliaksei, et al. "Active sampling for pairwise comparisons via approximate message passing and information gain maximization." *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021.

PQA – Detail preservation on realistic mannequins

Automated ROI extraction





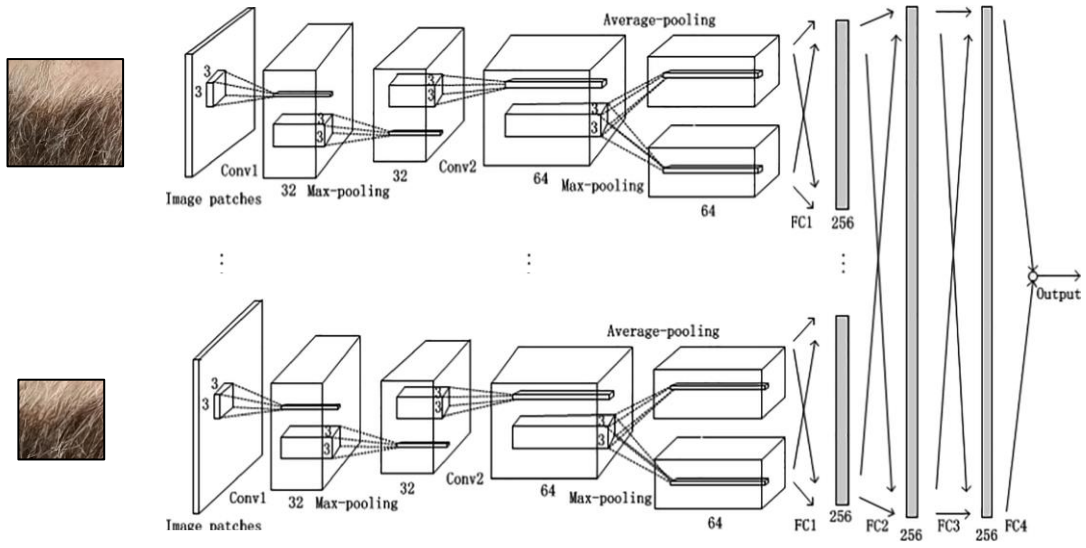


PQA – Detail preservation on realistic mannequins

Training approach:

- ROI images are divided into n random crops:
  - Fixed input size
  - Works as a data augmentation technique
  - Increases variance
  - Irrelevant crops

- Multi-Scale CNN
  - Multiple image scales: extract information on multiple levels
  - Fine tuning of MobileNetV3 pretrained on ImageNet
  - Augmentation using slight attribute alterations: noise, contrast, exposure, saturation and sharpening
  - Huber Loss to eliminate the effect of irrelevant crops



$$L_\delta = \begin{cases} \frac{1}{2}\left(y - f(x)\right)^2, & if\ |y - f(x)| \le \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & otherwise \end{cases}$$

Quadratic

Linear

PQA – Detail preservation on realistic mannequins

Evaluation:

- Pearson linear correlation coefficient (PLCC): $$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

- Spearman rank correlation order coefficient (SROCC): $$R = 1 - \frac{6\sum d^2}{n(n^2-1)}$$ Where $d = |R_1 - R_2|$

- Mean absolute error (MAE): $$\frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

PQA – Detail preservation on realistic mannequins

Lab setup limitations:

- Limited number of mannequins: hard to generalize on new mannequins and real faces
- Lab conditions do not generalize well to natural scenarios

Approaches:

- Develop a similar measure on real portraits
- Cover more random but controlled natural conditions

PQA – Detail preservation on realistic mannequins

Deep learning limitations:

- Random crops
- Limited generalization power
- limited dataset size, etc.

Approaches:

- Attention mechanisms
- Domain adaptation
- Self-supervision, etc.

PQA – Portraits in the wild

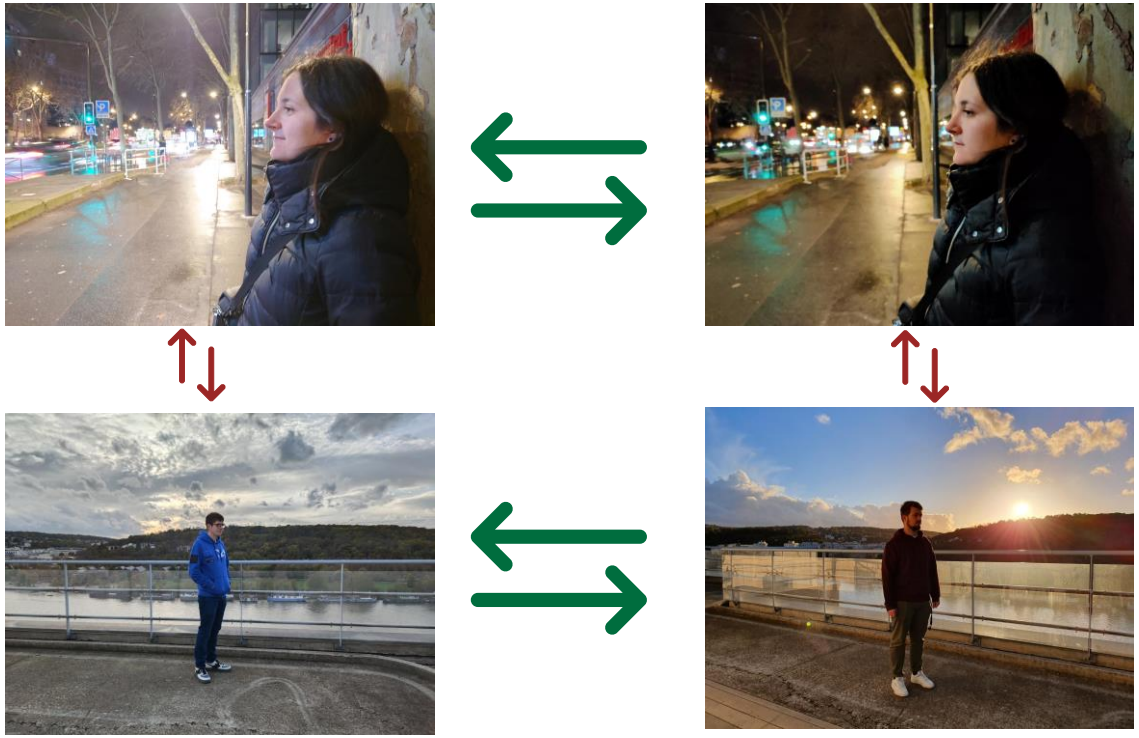
- Natural scenes are notoriously hard to evaluate, portraits in particular (camera variables, subject variables, environment variables, etc.)

- Many challenges occur:
  - How to generate a common quality scale for all portrait images?
  - How to conduct objective perceptual analysis?
  - What characteristics can be evaluated on natural portraits? How? Where?
  - How to handle ethnic and gender bias?
  - How to handle image rights?

PQA – Portraits in the wild – Challenges

How to generate a common quality scale for portrait images?

- Image quality is a relative metric: the more varying the content, the more subjective and noisier the common quality scale will become



Easy

Hard

PQA – Portraits in the wild – Challenges

How to generate a common quality scale for portrait images?

- Image quality is a relative metric: the more varying the content, the more subjective and noisier the common quality scale will become

- Common scales should be dissected and grouped, based on:
  - Lighting conditions
  - Skin tones
  - Weather conditions
  - Backgrounds
  - Framing
  - Reference pristine image
  - Etc.

PQA – Portraits in the wild – Challenges

To make the analysis as objective as possible, we can:

- Fix content (“scenes”): predefined visual content following specific protocols
- Define attributes: describe specific attributes to be analysed following strict expert guidelines
- Pairwise comparisons: with PWC we can prevent subjective unanchored quality evaluation by simplifying the task down to straightforward comparisons
- Regions of interest: set the zone to be analysed for each attribute as well as its resolution

PQA – Portraits in the wild – Challenges

What characteristics can be evaluated on natural portraits?

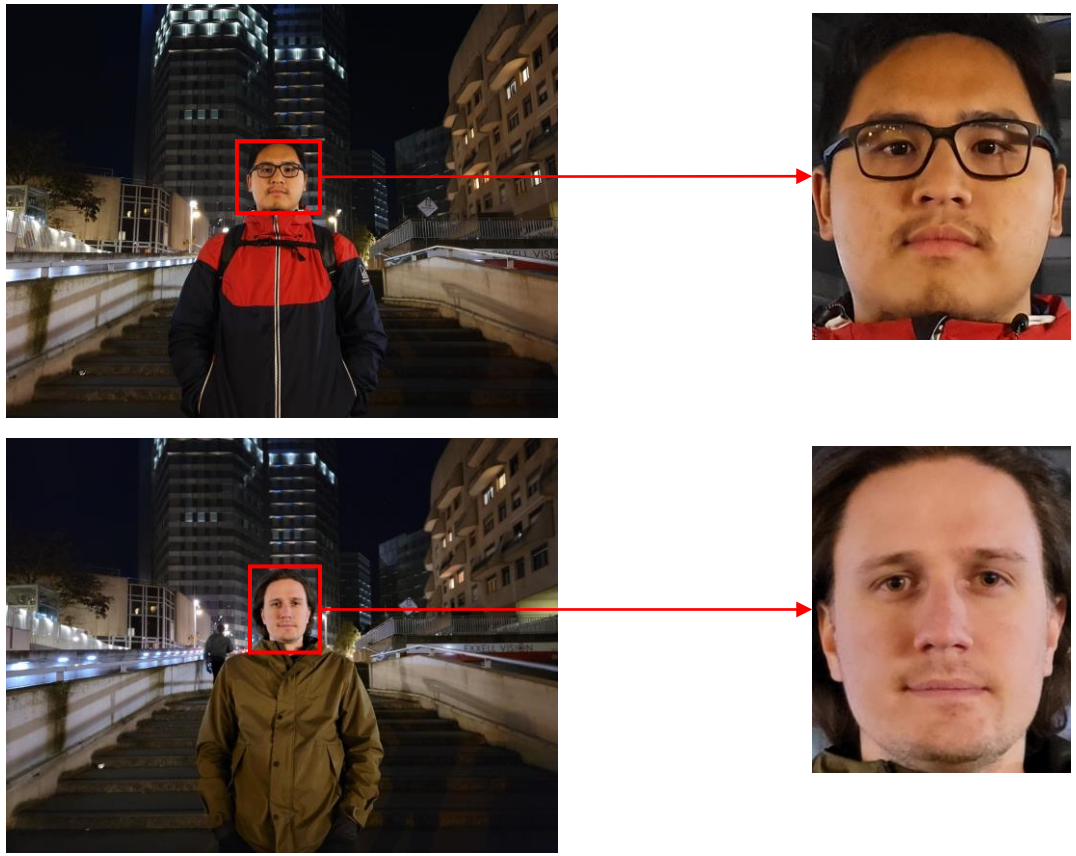
- In portraits, the rendering of the subject's features are of higher interest: Face details, face target exposure, skin smoothness, skin tone rendering, face shape distortions, red eyes, etc.

- Global attributes should also be considered: Overall target exposure, dynamic range, white balance (cast), color rendering, sharpness, focus range, depth of field, motion blur, etc.

- Artifacts should also be taken into consideration: Ghosting, halos, hue shift, ringing, flare, etc

PQA – Portraits in the wild – Challenges

Example: Detail preservation



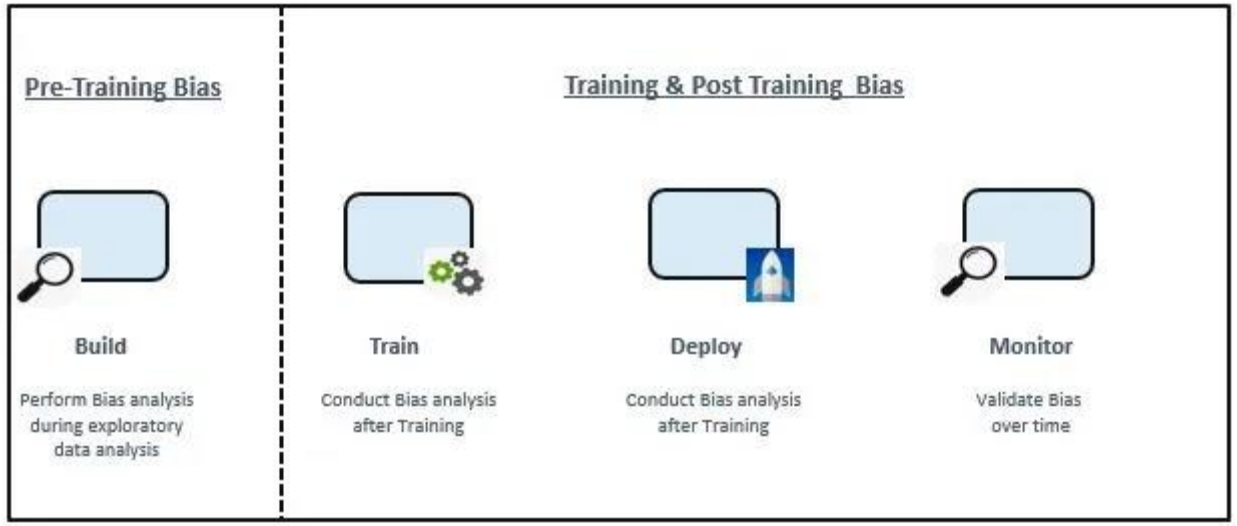
- Fixed aspect ratio
- Upscaled to the same resolution
- Images come from the same “scene”

PQA – Portraits in the wild – Challenges

How to handle ethnic and gender bias?

- Assure a decent coverage of subjects as well as annotators from different ethnicities and genders

- Don't hide it: assume the bias and find it through exhaustive testing (fairness Metrics, cross validation, elimination tests, etc.)



PQA – Portraits in the wild – Challenges

How to handle image rights?

- All the people appearing in the images should explicitly grant permission for image rights
- Tag each image with people's IDs
- Contractually require all data users to comply with relevant data protection laws, including the GDPR

PQA – Portraits in the wild – DL

What's the training approach?

- Each scene is defined by its own index, as a separate class
- Train on JOD scores and feed the class as input (training per scene)
- Random cropping
- Bootstrapping on scores to detect the variance in the annotations (more in the TP)
- Self-adaptive hypernetworks (HyperIQA [3])

[3] Su, Shaolin, et al. "Blindly assess image quality in the wild guided by a self-adaptive hyper network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Conclusions

Conclusions

1. In the last 5 years, we have seen a big improvement in smartphone camera technology thanks to AI

2. Current trends: « Photographer in your pocket »: set up parameters of the camera following the content on the scene as a photographer would do

3. Image quality assessment
   - Importance of protocol definition to avoid bias on the posed questions, knowledge of the annotator, etc which can render the results useless
   - critical for camera evaluation and (probably) for image curation
   - Big space for ML-based solutions

THANK YOU !

24-26, quai Alphonse le Gallo - 92100 Boulogne-Billancourt - France
www.dxomark.com